CYBERGIS-ENABLED REMOTE SENSING DATA ANALYTICS FOR DEEP LEARNING OF LANDSCAPE PATTERNS AND DYNAMICS

BY

ZEWEI XU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Geography
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Professor Shaowen Wang, Chair
Assistant Professor Kaiyu Guan
Assistant Professor Zhe Jiang, University of Alabama
Professor Bruce Rhoads
Professor E. Lynn Usery, U.S. Geological Survey

ABSTRACT

Mapping landscape patterns and dynamics is essential to various scientific domains and many practical applications. The availability of large-scale and high-resolution light detection and ranging (LiDAR) remote sensing data provides tremendous opportunities to unveil complex landscape patterns and better understand landscape dynamics from a 3D perspective. LiDAR data have been applied to diverse remote sensing applications where large-scale landscape mapping is among the most important topics. While researchers have used LiDAR for understanding landscape patterns and dynamics in many fields, to fully reap the benefits and potential of LiDAR is increasingly dependent on advanced cyberGIS and deep learning approaches. In this context, the central goal of this dissertation is to develop a suite of innovative cyberGIS-enabled deep-learning frameworks for combining LiDAR and optical remote sensing data to analyze landscape patterns and dynamics with four interrelated studies. The first study demonstrates a high-accuracy land-cover mapping method by integrating 3D information from LiDAR with multi-temporal remote sensing data using a 3D deep-learning model. The second study combines a point-based classification algorithm and an object-oriented change detection strategy for urban building change detection using deep learning. The third study develops a deep learning model for accurate hydrological streamline detection using LiDAR, which has paved a new way of harnessing LiDAR data to map landscape patterns and dynamics at unprecedented computational and spatiotemporal scales. The fourth study resolves computational challenges in handling remote sensing big data and deep learning of landscape feature extraction and classification through a cutting-edge cyberGIS approach.

To my family, friends, and colleagues with love

ACKNOWLEDGMENTS

This dissertation would not have been possible without the extraordinary guidance, help, and encouragement from my mentors, friends, and family. They make my PhD study at UIUC an exciting and memorable journey.

I would like to start by expressing my deepest gratitude to my advisor, Dr. Shaowen Wang. The past several years have not been an easy ride and I have learned numerous things from him, academically and spiritually. Throughout my time with him, he selflessly shares his wisdom of researching with me and constantly supports me when I hit nadirs in life. I still remember the numerous times when I struggled with research roadblocks and life unfortunates. It was he who taught me the importance of perseverance in doing research and encouraged me to get through my dark times. I would also like to thank Dr. Kaiyu Guan, Dr. Zhe Jiang for their patient and rigorous guidance. Their research enthusiasm and meticulousness encourage me throughout my journey with them. My appreciation also goes to Dr. Bruce Rhoads and Dr. E. Lynn Usery, who inspired me a lot during my PhD study and defense, and provided practical and valuable suggestions for my dissertation work.

Moreover, I would like to thank my best friends and colleagues in the Department of Geography and Geographic Information Science and Cyberinfrastructure and Geospatial Information Laboratory for their care, support, and companionship during the past six years. Many of these people have become my lifetime friends. Meanwhile, I want to give special thanks to Matthew Cohn, Susan Etter, and Miranda Czerwonka for their years of help and support.

Finally, I would like to thank my family for their unconditional love and support for my PhD study journey. I am extremely grateful to my father, Yuezhong Xu, who taught me the importance of self-discipline and positive thinking. And also my Mom, Xifeng Hu, who makes

me always believe in myself and also be kind and compassionate to others. There are always not enough words to express my appreciation for their sacrifices and love. Lastly, I am thankful for the patience of my future wife and unborn children, I will have more time to find and be with you from now on.

This dissertation research is based in part upon work supported by the U.S. National Science Foundation under grants: ACI-1443080, ICER-1833225, and OAC-1743184; and U.S. Geological Survey CESU Grant #G14AC00244. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation and the U.S. Geological Survey.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION AND BACKGROUND

1.1 Motivation and vision

Land cover and land uses greatly impact various climatic and land surface processes on Earth. These processes, including energy balance, carbon and hydrological cycles, and land-atmosphere interactions, depend on the physical and/or biogeochemical properties (e.g., albedo, emissivity, and photosynthetic capacity) of different land cover types (Foley et al., 2005; Brovkin et al., 2013; Bagley et al., 2014; Zhu and Woodcock, 2014; Costa et al., 2016). Land cover information has also been widely used by policy makers and practitioners for land management, environmental stewardship, and risk and disease controls (Homer et al., 2015). The accurate mapping of land cover is one of the most important tasks for the U.S. Geological Survey (USGS) in the past several decades (Homer et al., 2007).

Land use and land cover change detection is equally important compared to land cover mapping since it helps a broad spectrum of scientific, economic and governmental applications to understand the mechanism of human social development, project transportation and utility demand, identify future development pressure points and areas, and implement effective plans for regional development (Anderson, 1976). In recent years, considerably more attention is being directed towards monitoring changes in urban environments (Stow and Chen, 2002). Urban changes can affect various social, environmental, and economic conditions, including population migration, urban air quality, city infrastructure planning, business site locating, and urban greening. To better understand urban dynamics and urbanization often requires data-intensive change detection methods (Harrison and Hoyler, 2015). Data-driven characterization of urban building and building change information is beneficial to policy makers and city managers for

their decision making, but also important to urban habitants for better understanding their living environments (Coutard and Rutherford, 2015; Venerandi et al., 2017; Scott, 2017).

1.1.1 Landscape patterns and dynamics

The knowledge about land use and land cover has become increasingly important as the U.S. plans to overcome the problems of haphazard, uncontrolled development, deteriorating environmental quality, loss of prime agricultural lands, destruction of important wetlands, and loss of fish and wildlife habitat (McDermid et al., 2009; Wang et al., 2018; Aghsaei et al., 2020). Land use data are needed in the analysis of environmental processes and problems that must be understood if living conditions and standards are to be improved or maintained at current levels. One of the major prerequisites for better use of land is information on existing landscape patterns and changes through time. Knowledge of the present distribution and areas of agricultural, recreational, and urban lands, as well as information about their changing proportions, is needed by scientists, legislators, planners, and state and local governmental officials to determine optimal land use scenarios and policies, project transportation and utility demand, identify future development pressure points and areas, and implement effective plans for further development.

1.1.2 Artificial Intelligence for geospatial research

The recent advancement of Artificial Intelligence (AI) has provided tremendous opportunities for various scientific and application domains to tackle complex, and computation- and data-intensive problems (Russell and Norvig, 2016). As a branch of science that seeks to understand natural and human related phenomena according to location, geospatial research benefits from the recent development of AI by utilizing advanced machine learning algorithms

and analytical methods to extract important information from geospatial big data (VoPham et al., 2018; Hu et al., 2019a). There have already been a large number geospatial studies demonstrating the superiority of machine learning approaches to addressing research problems or questions that were previously difficult or impossible. In the domain of environmental health, for example, AI has been used to conduct accurate modelling of environmental exposures using geospatial data like Google Street View panorama images (Larkin and Hystad, 2019; Boulos et al., 2019). Also, AI has helped enabled autonomous vehicles and intelligent transport systems by incorporating a great amount of geospatial information gathered by traffic cameras and sensors (Toth and Paska, 2007; Toth et al., 2018; Hipps et al., 2017). As one of the most cutting-edge AI approaches, deep learning has been frequently applied to enable the extraction, understanding, and prediction of geospatial information including feature extraction from unstructured and geotagged text data across different languages (Hu et al., 2019b; Li et al., 2015; Paul et al., 2016), urban sprawl prediction (Ou et al., 2019; Liu et al., 2019; Poghosyan, 2018), study of diffusions models of disease and invasive species (Teng et al., 2018; Santosuosso and Papini, 2018; Wang et al., 2018), indoor navigation (Salamah et al., 2016; Bozkurt et al., 2015; Mo et al., 2016), and landscape feature extraction from remote sensing data (Xu et al., 2018; Xu et al., 2017; Valero et al., 2010).

1.1.3 Remote sensing data analytics and cyberGIS

During the past several decades, rapid development of remote sensing technologies have taken place, and as a result remote sensing big data have become widely available for scientific research and the general public. While a large quantity of large-scale optical data has been collected by earth orbiting satellites and aircrafts for decades, the advancement of airborne

LiDAR technology has enabled the collection of large-scale and high-accuracy LiDAR data in recent years. However, LiDAR data store densely distributed 3D points accompanied with a series of point attributes, requiring significant storage space and sophisticated techniques for data access and management. Also, the 3D structure of LiDAR makes related analysis much more complex and computationally intensive than optical data. For example, such data structures as octree or *kd*-tree are often needed for conducting geospatial analyses and queries of LiDAR data (Barber et al., 2008; Richter et al., 2013; Xu et al., 2015).

CyberGIS, defined as geographic information science and systems (GIS) based on advanced computing and cyberinfrastructure (Wang 2010; Wang 2016; Wang and Goodchild 2019), provides a desirable framework to resolve the analytical and computational challenges in the context of remote sensing big data by seamlessly integrating highly interactive spatial analysis tools, computationally intensive deep learning methods, and streamlined access to advanced cyberinfrastructure capabilities. CyberGIS holds great potential for overcoming the difficulties of large computational workloads of remote sensing data processing and analytics that exceed the capacity of conventional GIS approaches (Wang et al., 2013).

1.2 Research problems and questions

This dissertation research develops several deep learning models and cyberGIS capabilities for remote sensing data analytics applied to land cover classification, urban change detection, and hydrological streamline detection with the following questions addressed:

Land cover classification:

1. How to extract complex 3D features from LiDAR for land cover mapping?

2. How to use the full geometric and intensity information from LiDAR and effectively fuse such information with multi-temporal imagery for generating high-accuracy land cover maps?

Urban change detection:

1. How to conduct urban building change detection at individual building level using multi-temporal LiDAR?

2. How to classify different change types (e.g., demolition, construction) and quantify the volume of changes?

3. What is the difference of model performance in different urban environments (e.g., commercial, industrial, and residential areas)?

Hydrological streamline detection:

1. How to extract streamlines automatically using deep learning?

2. What are the pros and cons of using deep learning over traditional machine learning and flow accumulation methods?

CyberGIS-enabled remote sensing data analytics:

1. What is the performance of deep learning models when applied to larger geographic areas?

2. What are optimal ways of doing data processing and model training based on large reference datasets (>90,000 samples)?

3. What are the effects of imbalanced training data on modeling performance?

In order to answer these questions, the following research objectives have been pursued.

Chapter 2:

1. Investigate deep learning methods with a particular focus on a convolutional neural network algorithm for landscape feature extraction based on the distribution of 3D points from LiDAR.

2. Integrate LiDAR extracted features with multi-temporal optical data to generate high-accuracy land cover maps.

Chapter 3:

1. Develop an object-oriented classification strategy for change detection analysis at an individual building level.

2. Conduct spatial clustering and construct an alpha shape model for 3D building model construction and volume estimation, and classify different types of changes by forming building pairs and conduct volume and footprint change analysis.

3. Separately conduct accuracy evaluation over three types of building locations (e.g., commercial, industrial, and residential buildings).

Chapter 4:

1. Develop an attention U-net model for hydrological streamline extraction using LiDAR derived feature maps.

2. Compare the extraction results with traditional machine learning methods including SVM and ANN, and results generated from two flow accumulation models.

Chapter 5:

1. Develop a scalable land cover classification model by optimizing data processing and model training through the use of high-performance computing.

2. Conduct accuracy evaluation between imbalanced and balanced training scenarios, and compare the overall, user's, and producer's accuracies among different classes.

1.3 Remote sensing and deep learning methods

Remote sensing can be understood as a process of collecting data and information from interested targets by measuring the energy that is emitted or reflected from the targets (Campbell and Wynne, 2011). Remote sensing data can be collected by various carriers including satellites and aircrafts, and can be further categorized into active and passive data based on the two types of sensors. Passive remote sensing mainly includes data collected from optical sensors and represents the majority part of remote sensing data collected to date. During the past several decades, the rapid development of optical sensors and aerospace technologies has generated massive data and thus provided tremendous research opportunities in many research domains including for example environmental science, geomorphology, hydrology, urban planning, and social sciences (Reiche et al., 2016; Alvioli et al., 2018; Biancamaria et al., 2016; Albert et al., 2017; Bennett and Smith, 2017). On the other hand, active remote sensing has also experienced its rapid development since about two decades ago, and is capable of acquiring detailed information of the Earth surface through sensors with high accuracy and fine spatial and

temporal scales, and relatively low cost in the applications of forestry, landscape dynamics, archaeology, and atmospheric science (Liu et al., 2017; Devaney et al., 2015; Sinha et al., 2015; Rawat and Kumar, 2015; Golden et al., 2016; Guo et al., 2018).

1.3.1 Optical remote sensing

During the past several decades, optical remote sensing has provided an efficient way for large-scale monitoring of the Earth dynamics by measuring solar radiation reflection from the Earth's surface (Slater, 1980). Optical sensors are designed to measure a certain portion of the electromagnetic spectrum, which normally includes visible (400-700nm), near infrared (800-1100nm) and shortwave-infrared bands (1100-2500nm) (Asra, 1989). Since different targets would absorb and reflect energy differently at various wavelengths, this information is organized as optical imagery to capture their different spectral reflectance signatures. Based on the number of spectral bands, optical imagery can be categorized as panchromatic, multispectral, and hyperspectral imagery depending on the number of spectral bands used in the imaging process. Based on the spatial resolution, optical imagery can be classified as hyperspatial (<2m), medium (2-30m), and coarse resolution (>30m).

The high temporal resolution of optical remote sensing data has provided tremendous opportunities for various applications. Many studies in the context of optical remote sensing have demonstrated the effectiveness and feasibility of using multispectral satellite data for high accuracy land cover mapping (Kussul et al., 2017; Kim et al., 2018; Xu et al., 2018), environmental and socioeconomic monitoring (Shen et al., 2016; Bennett and Smith, 2017; Huang et al., 2017), change detection (Wang et al., 2015; Wang et al., 2017; Clement et al., 2018), and climatic or crop yield prediction (Firozjaei et al., 2018; Tuia et al., 2018; Zhou et al.,

2017; Singla et al., 2018). Among all the available sources of multitemporal data, the Landsat mission provides the archives of data since 1972 with near-global coverage (Wulder et al., 2016). Additionally, Landsat has relatively fine spatial resolution (30m pixel size) compared to other free data sources such as the Moderate Resolution Imaging Spectrometer (Xu et al., 2017).

1.3.2 LiDAR remote sensing

Conventional remote sensing methods based on optical data (e.g., visible and near-infrared bands) are capable of capturing horizontally distributed features, but are inherently limited in deriving sophisticated 3D information. Light detection and ranging (LiDAR) remote sensing, which utilizes near infrared light in the form of laser pulses to measure reflectance distance of targets (Schwarz, 2010). Laser pulses of LiDAR can penetrate many obstacles and provide multiple returns, which can generate precise, three-dimensional information of Earth surface characteristics. For example, small footprint airborne LiDAR can achieve the highest measurement accuracy of terrain features compared to other remote sensing modalities, even in wet regions or dense forests (Popescu et al., 2011). LiDAR remote sensing has now evolved as one of the most important tools to study landscape dynamics with its powerful capability to acquire extraordinarily high-accuracy 3D information at unprecedentedly high resolutions (Yan et al., 2015). While we are collecting a large amount of LiDAR data using advanced technologies, effective methods to extract critical and useful information from such data becomes equally important and increasingly challenging based on traditional remote sensing data analysis methods.

Extensive research has been done in using geometric components of LiDAR data to improve land cover classifications. Many researchers used LiDAR derived information, including digital surface model (DSM), digital terrain model (DTM), point density and spatial statistics

calculated from LiDAR data for distinguishing land cover types. Various classification methods are also used when including LiDAR data, such as maximum likelihood estimation (Bartels and Wei, 2006), object-oriented modeling (Brennan and Webster, 2006; Carlberg et al., 2009), neural networks (Nguyen et al., 2005), SVMs (Lodha et al., 2006), and other machine learning algorithms (Charaniya et al., 2004; Zhu and Toutin, 2011; Chen et al., 2017). Extensive findings suggest that statistical features derived from intensity values of LiDAR are also useful for distinguishing some classes that have little morphological variation (Brennan and Webster, 2006; Antonarakis et al., 2008; Bretar et al., 2008; Zhou, 2013; Morsy et al., 2016). More importantly, there has been an increasing interest in combining both optical data (e.g., multispectral imagery (Guo et al., 2011; Syed et al., 2005; Singh et al., 2012; Wulder et al., 2007; Lee and Jie, 2003), RGB imagery (Chen et al., 2009), high spatial resolution near-infrared imagery (Sasaki et al., 2012; Arroyo et al., 2010), and hyperspectral imagery (Dalponte et al., 2008) and LiDAR data to significantly improve the accuracy of land cover maps.

On the other hand, multitemporal LiDAR is superiorly important to capture urban dynamics in the vertical direction with a higher level of details and accuracy compared to optical data (Yan et al., 2015), as many urban changes take place vertically (e.g., building construction and demolition, infrastructure construction, and vegetation change). Extensive research has been done by transferring LiDAR points to rasters and conducting urban change analysis by calculating the difference between rasters (Teo and Shih, 2013). A better strategy is developed by constructing voxel grids to synthesize LiDAR points and detecting changes by checking the status of existence of points within voxels since changed voxels would have a change of status during multi-temporal periods (Xu et al., 2015). However, the relationship among points is often implicit and difficult to be fully represented using traditional 2D features (Du et al., 2016) and voxel-based features (Papon

et al., 2013; Maturana and Scherer, 2015a). These methods are not resilient to false and incomplete detection of changes because of misalignment of multitemporal datasets, inaccurate registration, moving objects issue, etc. (Xiao et al., 2015).

1.3.3 Deep learning

Deep learning, which is also called deep structured learning or hierarchical learning, represents a class of algorithms that conduct supervised learning in a hierarchical way from different levels of feature abstraction (Goodfellow et al., 2016). Different statistical models are incorporated in deep learning algorithms for extracting knowledge and constructing concepts from lower to higher levels, which at the same time reduce redundant information to the largest possible degree (LeCun et al., 2015; Schmidhuber, 2015). As the most important advancement of artificial intelligence during the past decade, deep learning still has a fast-growing pace and has been proven to be applicable to many domains of science, business and government (Zhu et al., 2017).

Deep learning algorithms usually consist of neural networks with more than two hidden layers in contrast to traditional Artificial Neural networks (ANNs) (Deng and Yu., 2014). The biggest advantage of deep learning over traditional ANNs is that it eliminates the traditional feature-crafting process that requires specific domain knowledge, by directly using raw data as input for different applications, and achieves better performance through hierarchical feature learning. In recent years, deep learning research has been extensively pursued by numerous research agencies and also large industrial cooperation, and it has dominated various research areas that rely on machine learning including image recognition, motion planning in autonomous systems, speech recognition, diseases diagnosis and prediction, biomedical analysis, natural

language processing, and so on. Deep learning is expected to continue to advance with the development of more new architectures and have many more successes in the near future by requiring minimal engineering efforts, and taking advantage of increases in the amount of available computational power and big data (LeCun et al., 2015).

The emergence of deep learning technologies has also paved a new way for remote sensing data analysis. The low-level features (e.g., spectral, texture, and geospatial information) from remote sensing data can be directly fed into the deep nets for transformation into higher level of feature representations for further analysis. Based on the different output requirements of specific remote sensing questions, deep learning algorithms can be adapted at layer-level in flexible ways. Currently, deep learning algorithms have been used in various remote sensing data analysis: from the traditional topics of image preprocessing, pixel-based classification, and target recognition, to the recent challenging tasks of high-level semantic feature extraction and remote sensing scene understanding (Zhang et al., 2016). Among a large variety of deep learning algorithms, one of the most successful types is called Convolutional Neural Networks (CNNs). During the training process, CNNs learn hierarchical features by applying convolutional filters, and unimportant information is gradually reduced during the convolutional processes. CNNs have proven to be effective in various remote sensing image classification tasks (Maggiori et al., 2017; Nogueira et al., 2017; Xu et al., 2018), and also tasks such as ground object recognition (Kampffmeyer et al., 2016; Long et al., 2017) and high-resolution aerial image segmentation (Sun et al., 2018).

The recent development of 3D deep learning algorithms provides researchers new opportunities to solve problems from 3D perspectives. For example, depth image is a source for the third dimension that is combined with normal RGB image to form RGBD for 3D CNN

applications including 3D object recognition (Gupta et al., 2014; Wu et al., 2015; Alexandre, 2016), and semantic segmentation (Hoft et al., 2014). The work by Prokhorov (2010) developed a 3D CNN with LiDAR data applied to a binary object classification problem. After that, Maturana and Scherer (2015a, 2015b) designed a generalized 3D CNN method called Voxnet for object classification using the full volumetric point cloud information from LiDAR, which was proven to be superior to other 3D CNNs. They also pointed out the future of using intensity attributes for classifying more complex scenes (Maturana and Scherer, 2015b).

In order to preserve point cloud information to the largest extent, several point-based deep learning models have been proposed. The pioneer one is called PointNet (Qi et al., 2017a), which takes raw point cloud as input. It not only accelerates the computation but also notably improves the accuracy of many point-based classification tasks (Qi et al., 2017a; Yousefhussien et al., 2018). Pointnet++ is an advanced version of Pointnet, which incorporates a hierarchical structure of point neighborhood learning through points downsampling and interpolation (Qi et al., 2017b). However, both Pointnet and Pointnet++ are limited in their ability to capture complex shape patterns based on a simple design of orientation-encoding (*k*-nearest neighbor searching) and less scale awareness for feature calculation (Jiang et al., 2018). The scale-invariant feature transform (SIFT) excels over many other feature encoding methods by using a strategy of multi-orientation feature encoding and a scale-aware design (Furuya and Ohbuchi, 2009; Darom and Keller, 2012). PointSIFT, which incorporates this idea to treat 3D point clouds, has shown its robustness by outperforming the state-of-the-art methods including Pointnet and Pointnet++ on S3DIS (Armeni et al., 2016) and ScanNet (Cohen et al., 2018) datasets (Jiang et al., 2018). The PointSIFT architecture incorporates 8-direction orientation-encoding (OE) units from multiple spatial scales into deep neural networks. In this way, the neurons in different

stacked OE units can perceive different scales. The scheme is to put these OE units together by shortcut connections and let neural network (after training) select the appropriate scale. Like Pointnet++, PointSIFT also utilizes the two-stage (i.e., encoding (downsampling) and decoding (upsampling)) for feature learning, which significantly improves the presentation capacity of the network.

1.4 Structure of the dissertation

This dissertation contains six chapters in total. Chapter 1 is an overview chapter that introduces the background, motivation, and contributions of the research. From Chapter 2 to Chapter 5, each chapter is an individual piece of research focusing on different research questions posed in Section 1.2 and is a peer-reviewed article under the condition of published, under review, or to be submitted.

Chapter 2 develops a novel deep learning framework for land cover classification by utilizing LiDAR and multitemporal Landsat images. Landscape has complex 3D structures. Previous research conducted analysis solely using traditional LiDAR metrics (e.g., variance, entropy, skewness, kurtosis). There has been little work on using deep learning feature extraction to help understand and extract features by harnessing the full 3D point cloud and intensity data that LiDAR collects. In this research, we adapt a 3D deep learning architecture specifically to solve the problem for land cover classification using both LiDAR and imagery data. During the training, the fully connected layer in the last epoch is removed and the last dense layer is extracted as the output. Then these features are combined with multitemporal spectral data for the final classification using a multi-class SVM classifier. We are also the first to develop the intensity grid for more complete LiDAR feature learning than previous studies (Brennan and

Webster, 2006; Zhou, 2013). The major contributions of this research are two-fold: (1) a novel model for land cover classification by using full geometric and intensity information extracted from LiDAR data and multitemporal imagery, and (2) an effective 3D buffering method for resolving edge issues in the training data augmentation process while achieving high accuracy by using a post-voting strategy.

Chapter 3 shifts the subject to urban change detection. Urbanization has intensified across the globe at an unprecedented pace during the past several decades. Consequently, the demand for accurately acquiring urban building and building change information is expected to continue its increasing trend in the foreseeable future. LiDAR data contain full 3D information in the form of point clouds and are especially advantageous in detecting urban building changes. Previous research using LiDAR was significantly restricted by its rigid requirement of registration rate, rule-based removal of non-building objects, and high sensitivity due to comparison with mixed classes of points. In this research, we use a deep learning method to conduct a point-based urban building classification by adapting the PointSIFT algorithm to airborne LiDAR. Then we use an Euclidean-distance based clustering method to separate individual buildings and estimate building footprints. Finally, we infer building changes and classify change types by estimating the change of building volume and footprint from 3D alpha shapes, which is a generalization of the 3D convex hull and a subgraph of the Delaunay triangulation over a set of 3D points. Four types of changes including demolition, new construction, reduction, and expansion are extracted based on their different patterns. A reference dataset which is created by visual interpretation from both a multitemporal aerial photography and LiDAR data are used to separately evaluate the recall (completeness), precision (correctness), and F1 score of building classification and

change detection results. The estimated volumes in 2014 are also validated from 3D building survey data from the government data portal of Boston.

Chapter 4 investigates an important topic in hydrography-streamline detection. The accurate delineation of hydrological streamlines, which is one of the major forms of land surface water, is critically important in various scientific disciplines, such as the assessment of present and future water resources, climate models, agriculture suitability, river dynamics, wetland inventory, watershed analysis, surface water survey and management, flood mapping, and environmental monitoring. Traditional hydrological models generate streamlines solely based on topological information, which inevitably contain errors. For example, dried out drainage lines would always be falsely recognized as streamlines. Traditional methods also ignore the information from the complex 3D environment of streamlines and surface reflectance information, which would potentially be helpful to accurately delineate streamlines. In recent years, the availability of high accuracy LiDAR data provides us a promising method to capture both 3D information of the environment and also surface reflectance information of land cover. LiDAR sensor uses NIR light in the form of a pulsed laser to measure ranges (variable distances) to the ground and also reflectance information at multiple returns. These light pulses generate precise, three-dimensional information about the shape of the surface characteristics. In this research, multiple LiDAR features maps are generated, and we developed an attention U-net model for doing the streamline detection and we also tested several traditional machine learning methods as our baseline for comparison.

Chapter 5 focuses on the computational scalability of the workflow for land cover classification using LiDAR and multitemporal Landsat imagery and expands the research area to 34 counties of southeast Illinois. Although LiDAR data contain rich information about 3D

characteristics of various landscape types, the 3D feature extraction of a large amount of LiDAR data are often challenging mainly due to the complexity of such data and computational intensity. This research mainly tackles the computational challenges by utilizing cyberGIS-enabled cyberinfrastructure. To be specific, we utilize high throughput computing for LiDAR data preprocessing using 17 computing nodes (200+ CPU cores). Data reprojection and noise removal functions are applied in parallel for each tile of LiDAR. We also optimize the LiDAR sample points queries using spatial indexing in the process of training, validation, and prediction sample preparation. After that, we conduct data augmentation and voxelization in parallel for the data samples and use the GPU resource from Chameleon cloud service for a 3D CNN model training. In the classification stage, we optimize the RBF kernel calculation by using a GPU for the classification of seven land cover types. As the last step, we utilize a multi-GPU architecture to do model prediction over the research area to generate the final classified map.

Chapter 6 is the last chapter that summarizes the contributions of the dissertation, describes the limitations, and points out future work and potential research directions.

1.5 Synthesis and contribution

Most of current deep learning models in remote sensing data analysis are designed based on 2D learning or with limited expression of 3D information at a small scale. There are limited progress on 3D deep learning frameworks that are specifically designed for analyzing geospatial big data in 3D. The major contribution of this research is to develop and apply 3D deep learning frameworks to 3D geospatial big data in terms of landscape feature extraction and change detection. The primary goal of this research is to establish a cyberGIS-enabled deep learning framework to perform landscape feature extraction and classification using LiDAR data and

optical remote sensing data. The proposed methods in this framework not only pave a new way for understanding landscape patterns and dynamics in 3D, but also demonstrate its large-scale applications. The major conclusions from the four interrelated research investigations highlight the importance of advancing cyberGIS-enabled deep learning for mapping land cover patterns and dynamics using remote sensing big data.

1.6 Publications

Chapter 2: Xu, Z., Guan, K., Casler, N., Peng, B., & Wang, S. (2018). A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal Landsat imagery. ISPRS journal of photogrammetry and remote sensing, 144, 423-434.

Chapter 4: Xu, Z., Guan, K., & Wang, S. (under review). Urban building extraction and change detection using multitemporal LiDAR data based on a deep learning and rule-based method. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Chapter 5: Xu, Z., Wang, S., Stanislawski, L. V., Jiang Z., Jaroenchai, N., Sainju, A. M., Shavers, E., Usery E. L., Chen L., Li, Z., & Su, B. (under review). Hydrological streamline detection using a U-net model. Nature Machine Intelligence.

Rahil,S., Xu, Z., Sugumaran, R., Oliveira, S.(2016). Parallel Landscape Driven Data Reduction Spatial Interpolation Algorithm for Big LiDAR Data. ISPRS International Journal of Geo-Information, 5, no.6:97.

Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., Xu, L., Chen, Z., Gong, J. (2019). A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. Environmental Modelling & Software, 104502.

Cienciala, P., Nelson, A. D., Haas, A. D., & Xu, Z.. (2020). Lateral geomorphic connectivity in a fluvial landscape system: Unraveling the role of confinement, biogeomorphic interactions, and glacial legacies. Geomorphology, 107036.

Chen, L., Xu Z., Li H., Wang, S. (under review). Should we trust the deep neural network for remote sensing image classification? A Comprehensive Analysis. ISPRS Journal of Photogrammetry and Remote Sensing.

Chen, L., Xu Z., Li H., Wang, S. (under review). DAPnet: A double attention convolutional network for segmentation of point clouds. ISPRS Journal of Photogrammetry and Remote Sensing.

Lyu, F., Xu Z., Wang, S., Li Z., Ma, X., Wang, S. (under review). A Vector-Based Method for Drainage Network Analysis Based on LiDAR Data. Computers and Geosciences.

CHAPTER 2: A 3D CONVOLUTIONAL NEURAL NETWORK METHOD FOR LAND COVER CLASSIFICATION USING LIDAR AND MULTI-TEMPORAL LANDSAT IMAGERY

Abstract

The terrestrial landscape has complex three-dimensional (3D) features that are difficult to extract using traditional methods based on 2D representations. These methods often relegate such features to raster or metric-based (two-dimensional) representations based on Digital Surface Models (DSM) or Digital Elevation Models (DEM), and thus are not suitable for resolving morphological and intensity features for fine-scale land cover mapping. Small-footprint LiDAR provides an ideal way for capturing these 3D features. This research develops a novel method of integrating airborne LiDAR derived features and multi-temporal Landsat images to classify land cover types. We tested our approach in Williamson County, Illinois, which has diverse and mixed landscape features. Specifically, our method applied a 3D convolutional neural network (CNN) method to extract features from LiDAR point clouds by (1) creating an occupancy grid, an intensity grid at 1-meter resolution, and then (2) normalizing and incorporating data into the 3D CNN. The extracted features (e.g., morphological and intensity features) from the 3D CNN were finally combined with multi-temporal spectral data to enhance the performance of land cover classification based on a Support Vector Machine classifier. Visual interpretation from both hyper-resolution photos and point clouds was used for training and preparation of testing data. The classification results show that our method outperforms a traditional method by 2.65% (from 81.52% to 84.17%) when solely using LiDAR and 2.19% (from 90.20% to 92.57%) when combining all available images. We demonstrate that our

method can effectively extract LiDAR features and improve fine-scale land cover mapping through fusion of complementary types of remote sensing data.

2.1 Introduction and background

Land cover and land uses greatly impact various climatic and land surface processes. These processes, including energy balance, carbon and hydrological cycles, and land-atmosphere interactions, depend on the physical and/or biogeochemical properties (e.g., albedo, emissivity, photosynthetic capacity) of different land cover types (Foley et al., 2005; Brovkin et al., 2013; Bagley et al., 2014; Zhu and Woodcock, 2014; Costa et al., 2016). Land cover information has also been widely used by policy makers for land management, environmental stewardship, and risk and disease controls (Homer et al., 2015).

Satellite remote sensing has been widely used as an effective and efficient means to monitor land cover patterns at a large geographic extent. However, traditional remote sensing methods based on optical data (i.e. visible and near-infrared) are only suitable for capturing horizontally distributed features including shapes, structures, and areas. Because of the complexity of land cover types, Light Detection and Ranging (LiDAR) data becomes more frequently used to provide more information about land cover and enable the use of 3D characteristics of various landscape types to generate high quality land cover maps. Much research has been done in using geometric components of LiDAR data to improve land cover classifications. Many researchers used LiDAR derived information, including digital surface models (DSM), digital terrain models (DTM), and point density and spatial statistics calculated from LiDAR data for distinguishing land cover types. Various classification methods are also used when including LiDAR data, such as maximum likelihood algorithms (Bartels and Wei, 2006), object-oriented modeling (Brennan and

Webster, 2006; Carlberg et al., 2009), neural networks (Nguyen et al., 2005), SVMs (Lodha et al., 2006), and other machine learning algorithms (Charaniya et al., 2004; Zhu and Toutin, 2011; Chen et al., 2017). Extensive findings suggest that statistical features derived from intensity value of LiDAR are also helpful in distinguishing some classes that have little morphological variation (Brennan and Webster, 2006; Antonarakis et al., 2008; Bretar et al., 2008; Zhou, 2013; Morsy et al., 2016). More importantly, there has been an increasing interest in combining both optical data including multispectral imagery (Guo et al., 2011; Syed et al., 2005; Singh et al., 2012; Wulder et al., 2007; Lee and Jie, 2003), RGB imagery (Chen et al., 2009), high spatial resolution near-infrared imagery (Sasaki et al., 2012; Arroyo et al., 2010), hyperspectral imagery (Dalponte et al., 2008) and LiDAR data to achieve improved accuracy of land cover maps. In these processes, LiDAR derived metrics and spectral bands are combined and fit into various algorithms for classification. These algorithms include random forest (Guo et al., 2011), Support Vector Machines (SVMs) (Bretar et al., 2008; García-Gutiérrez et al., 2015), object-oriented classifiers (Syed et al., 2005; Sasaki et al., 2012; Arroyo, 2010; Chen et al., 2009), rule-based model (Huang et al., 2008), and structural and intensity surface models (Singh et al., 2012). While these analyses have provided key insights into high-resolution land cover mapping, they all conducted analysis based on traditional LiDAR metrics (e.g., variance, entropy, skewness, kurtosis), and there has been little work on using deep learning feature extraction framework to help understand and extract features at a much more complex and abstract level by using the full three-dimensional point cloud and intensity data that LiDAR collects.

During recent years, biologically inspired Convolutional Neural Networks (CNNs) have emerged and proven to be effective in various pattern recognition and object classification tasks (Krizhevsky et al., 2012; Razavian et al., 2014). During the training process, CNNs learn

hierarchical features, corresponding to different levels of abstraction. Novel architecture of CNNs have been widely acknowledged as the most successful deep learning approach, and used as dominant methods in many recognition and detection tasks (Krizhevsky et al., 2012; Simonyan and Zisserman, 2015; Jia et al., 2014; Razavian et al., 2014; Oquab et al., 2014). However, the majority of CNNs are constructed based on 2D layers, which makes it difficult in solving classification tasks in 3D contexts. The recent development of 3D CNNs provides researchers new opportunities to solve classification problems in 3D space. Researchers can use different information for the third dimension. Temporal information is commonly used in video-based human action recognition (Ji et al., 2013). Some data have unique 3D information that is used for object recognition, such as tomography imagery with height information (Flitton et al., 2012), medical profile images from three dimensions as inputs (Kamnitsas et al., 2017). Hyper-spatial, multi-band imagery, which has different bands as the third dimension, is also used for 3D CNN model training in several land cover classification studies (Chen et al., 2016; Audebert et al., 2017; Ji et al., 2018). A depth image is another source for the third dimension that is combined with normal RGB images to form RGBD for 3D CNN applications including 3D object recognition (Socher et al., 2012; Gupta et al., 2014; Wu et al., 2015; Alexandre, 2016), and semantic segmentation (Hoft et al., 2014). The work by Prokhorov (2010) developed a 3D CNN with LiDAR data applied to a binary object classification problem. After that, Maturana and Scherer (2015a, 2015b) designed a generalized 3D CNN method called Voxnet for object classification using the full volumetric point cloud information from LiDAR, which was proven to be superior to other 3D CNNs. They also pointed out the future of using intensity attributes for classifying more complex scenes (Maturana and Scherer, 2015b). In addition, voxelization of intensity data is a solid way for data normalization (Reymann and Lacroix, 2015).

In this research, we have developed a novel method for complex feature learning from LiDAR data to improve land cover classification. The Voxnet architecture has several advantages that make it suitable for our research problem. First, the architecture is able to learn local spatial filters useful to the classification task. In our case, we use activation maps from filters to encode structures such as planes and corners at different scales. Second, the hierarchy of more complex features can be extracted based on the stacking of multiple CNN layers over the entire feature space. Third, our new method is trained fully from the raw volumetric data based on the activations of each sub volume of data, where probabilities are finally calculated for class prediction. However, Voxnet is a 3D CNN classifier, which only allows the input of single type of data. Thus we adapt the architecture specifically to solve our problem for classification using both LiDAR and imagery data. During the training, the fully connected layer in the last epoch is removed and the last dense layer is extracted as the output. Then these features are combined with multitemporal spectral data for the final classification using a multi-class SVM classifier. The major contributions of this research are two-fold: (1) a novel method for land cover feature extraction by using full geometric and intensity information from LiDAR data, and (2) an effective data fusion model of using 3D features from LiDAR and multi-temporal images to achieve high accuracy in land cover classification.

2.2 Study area and dataset

2.2.1 Study area

Our study area includes the majority of Williamson County, Illinois, USA (Figure 2-1). This area has 1,138 km^2 with a population of 66,357. It has a mix of humid continental and subtropical climatic types with four distinct seasons. Spring is the wettest with extreme weather

including tornadoes and winter storms. Summer is humid with warm temperature sometimes reaching 100 °F (38 °C). Fall is mild with heavy rainfall. Winter has periodic snow when the temperature reaches around 32 °F (0 °C). This area contains a complex composition and heterogeneous distribution of forest, shrub land, wetland, grassland and various types of agricultural land, water body, and developed land.

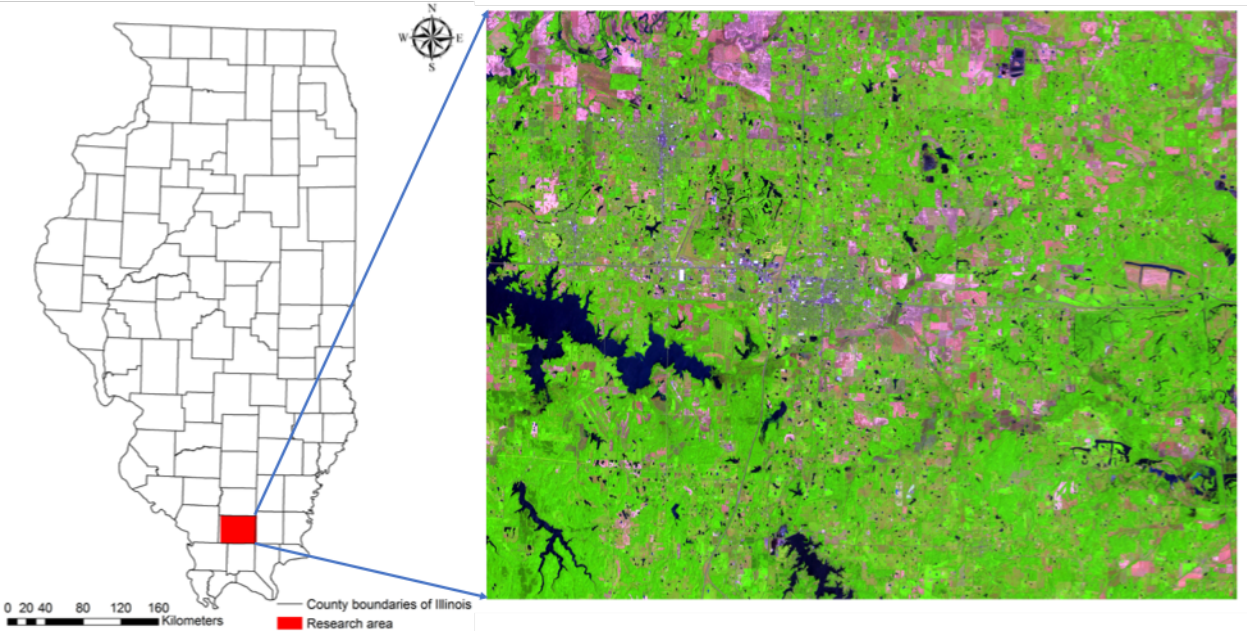


Figure 2-1. Research area (Williamson County, Illinois; Landsat 5 TM image: bands composition 5, 4, 3 and taken on July 22, 2011)

2.2.2 Dataset

Regarding the spectral data side, this study uses twelve cloud-free Landsat 5 surface reflectance scenes downloaded from the USGS Earth Explorer Data Portal (https://earthexplorer.usgs.gov/). These images were separately acquired on August 20, August 27, September 5, September 12, September 21, October 7, October 14, October 30, 2010 and

January 2, April 17, June 4, July 22, 2011. Six TM bands (blue, green, red, near IR, and two mid IR bands) from each image are utilized to provide the spectral information for the classifications.

520 tiles of LiDAR data were flown from April 6-17, 2011 covering the research region and downloaded from the Illinois Geospatial Clearing House (https://clearinghouse.isgs.illinois.edu). The data were collected by Federal Emergency Management Agency and rectified by multiple companies to ensure both vertical and horizontal accuracy. The vertical accuracy is 8.53 cm RMSE and horizontal accuracy of 0.2 meter with a point spacing of 1.4 meters. The instrument collected up to 4 discrete returns per pulse, with intensity readings of 12-bit dynamic range per measurement, at around 1045nm. The delivered data had an average density of 2 points per m^2 and ranged from 1-5 points per m^2 among different tiles.

The Illinois Department of Transportation (IDOT) Orthophotography that was flown in April, 2011, using a Z/I Imaging Digital Mapping Camera (DMC) with RGB at a resolution of 1 feet, was downloaded from the Illinois Geospatial Clearing House (https://clearinghouse.isgs.illinois.edu) and used as the source to generate reference data through visual interpretation. Because the twelve images, LiDAR, and reference source are all within a 12-month period, the land cover is assumed to be consistent and thus the same reference data can be used for classification of all images and LiDAR.

2.3 Methods

2.3.1 Method overview

The new method consists of data preparation, 3D CNN-based LiDAR feature extraction and classification. The feature extractor is adapted from Voxnet, which contains four types of

layers: convolutional, pooling, dropout and fully connected. The primary purpose of the convolutional layer is to extract features from original images. The convolutional process preserves the spatial relationship between pixels by learning image features using predefined multidirectional kernels to filter small squares of input data. The dropout layer is designed to randomly drop out neurons in order to avoid overfitting. The pooling layer conducts spatial down sampling by reducing the dimensionality of each feature map while retaining important spatial information. The fully connected layer is a traditional multi-layer perceptron that uses a softmax function to produce categorical probability distribution of a certain number of classes to the output layer. The fully connected layer conducts different classification tasks based on these features and training labels. In our study, we revised Voxnet as a feature extractor by abandoning fully connected layers during the last training epoch. Also, the standard input of Voxnet is cubic data volume. We instead use rectangular cuboid ($30\times30\times50$ m^3), where the first and second dimensions correspond to the resolution of the spectral data and the third dimension corresponds to height dimension. In the training process, we adjusted the batch size to fit our memory and set the maximum training epochs to 800.

2.3.2 Sampling design

In order to collect representative samples of each class in the population, it is important to keep the inclusion probability the same for each pixel (Xu et al., 2017). Thus, we selected 9000 pixels randomly on a 999×1282 fishnet grid (30 m) based on the Landsat imagery and overlaid on the orthophotography. On-screen interpretation of both LiDAR and photography and digitization are utilized to generate reference data. Finally, 8297 of them are useful and 703 are abandoned either because of cloud coverage or image registration errors. The general

interpretation principle is that a sample is assigned a class label if that class has a dominant coverage (>50%). One exception is that any sample contains more than 30% of developed type would be classified as developed to consider the connectivity of roads. Explicitly, we define herbaceous as dominant vegetation coverage lower than 1 meter; shrub as dominant vegetation coverage higher than or equal to 1 meter and lower than 5 meters; forest as dominant vegetation coverage higher than or equal to 5 meters. Both LiDAR and orthophotography were flown in April, which is the most humid season in the research area, so wetlands are defined as any vegetated area that is saturated with water. Among the 8297 samples, 70% (5808) are adopted as training data and 30% (2489) are adopted as testing for accuracy evaluation. Details of the reference data are listed in Table 2-1.

Table 2-1. Training and testing data number

	Agriculture	Developed	Wetlands	Forest	Herbaceous	Shrubland	Water	Total
Training	971	525	411	2191	810	541	359	5808
Testing	416	225	176	939	347	232	154	2489

2.3.3 Data preprocessing

Twelve multi-temporal land surface reflectance images are georeferenced with orthophotography and reprojected from WGS84/UTM 16N to NAD83/UTM 16N. They are also cropped based on the boundary of Williamson County using ArcGIS 10.4 data management toolbox (ESRI, 2016). Six bands (blue, green, red, near IR, and two shortwave IR bands) from each image are extracted, and the normalized difference vegetation index (NDVI) (Carlson and Ripley, 1997) is calculated for each image using red and near-infrared bands of the TM imagery. Additionally, five texture-based statistics are calculated using grey level co-occurrence matrices (GLCMs) of each image. These layers are stacked as the spectral data for classification. The

texture features include contrast, dissimilarity, homogeneity, energy and correlation and their formulas are shown below (Chen et al., 1998).

$$Contrast = \sum_i \sum_j (i-j)^2 P_d(i,j) \quad (2.1)$$

$$Dissimilarity = \sum_i \sum_j P_d(i,j)|i-j| \quad (2.2)$$

$$Homogeneity = \sum_i \sum_j \frac{P_d(i,j)}{1+|i-j|} \quad (2.3)$$

$$Energy = \sum_i \sum_j P_d(i,j)^2 \quad (2.4)$$

$$Correlation = \sum_i \sum_j \left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right] \quad (2.5)$$

where P_d is the co-occurrence matrix in a rectangle neighborhood (7x7) of a center pixel, i is the row number and j is the column number in P_d. σ^2 and μ are separately the variance and mean of a row or column of values in P_d.

Point data abstraction library (Pdal), a C++ based open source point cloud processing library, is used for LiDAR data preprocessing. LiDAR data are initially reprojected from the State Plane Illinois East US feet (2011) to NAD83 UTM 16N to keep it consistent with spectral data. A statistics-based outlier filter was used to get rid of isolated noisy points including birds, powerlines, etc. We first calculated mean distance of each point to its twelve neighbors. Any point falls below or above two standard deviations of the mean distance of all the points in 3D space would be treated as an outlier and removed. The pipeline of LiDAR data preprocessing is shown in Figure 2-2 below.

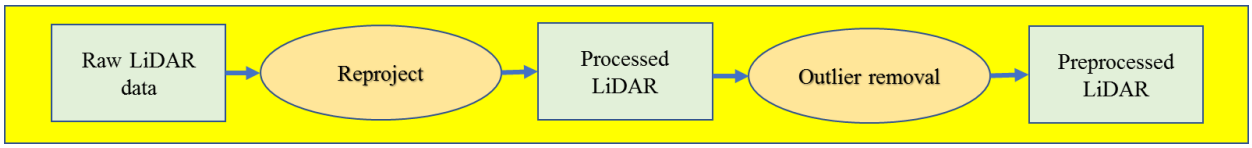


Figure 2-2. Pipeline of LiDAR data preprocessing

2.3.4 Benchmark method

The benchmark method consists of the following three components: data preparation, 2D-CNN-based LiDAR feature extraction, and classification. Data preparation includes ground/non-ground classification, generation of feature maps from ground/non-ground returns, and LiDAR statistics calculation. In this process, the raw LiDAR data are reclassified into ground and non-ground classes using LAStools (Isenburg, 2011). Then ground/non-ground points are separately used to generate DSM, DEM, ground and non-ground intensity maps at a resolution of 1m using blast2dem function in LAStools (Isenburg, 2011). The four feature maps are then stacked and feature map patches with the dimension of 32×32×4 are extracted for each sample. Five types of LiDAR statistical features (kurtosis, variance, coefficient of variation, skewness, and entropy) are separately calculated from the elevation and intensity distribution of points of each sample. In the second stage, a 2D CNN model is constructed as shown in Figure 2-3. This model consists of three convolutional layers with 32 filters at the size of (3, 3) and a step of (1,1), three maximum pooling layers with a step 2, a dropout layer with dropping rate of 0.45, a fully connected layer, and an output layer. The model is trained 800 epochs, and the last dense vector is extracted after the training is done. Different dimensions of dense vector (32, 64, and 128) are tested and 32 is finally selected because of its superiority. In the classification stage, the five types of statistical features are combined with the dense vector extracted from the 2D CNN model and LSR data for the final classification using a multi-class SVM classifier. The parameter tuning is conducted based on a three-level grid search and ten-fold cross-validation strategy. The best combination of parameters is used on the testing data for accuracy evaluation.

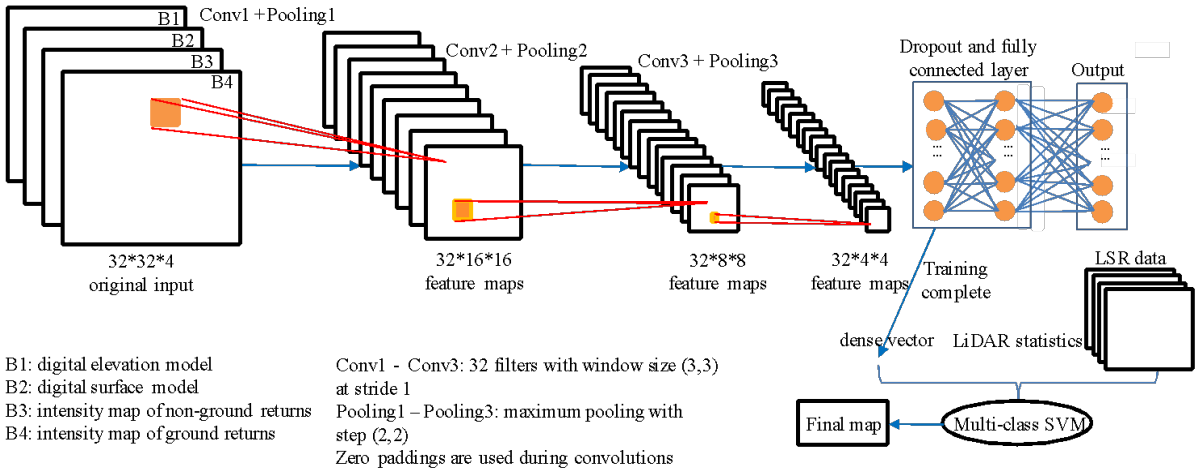


Figure 2-3. Flow chart of a 2D-CNN-based benchmark method

A standardized processing chain for DSM generation is utilized to map the above-ground height information in the study area using LAStools (Isenburg, 2011) and Pdal (https://www.pdal.io). To avoid edge artifacts in the following image processing procedures, non-overlapping LiDAR tiles are retiled with a 60-meter buffer around each of them. First, LiDAR points are classified into ground points and non-ground points, and the digital terrain model is generated with the ground points using the triangular irregular network (TIN) (Isenburg et al., 2006). Raw LiDAR points are then height-normalized by subtracting the digital terrain model from the orthometric heights at the corresponding locations. Finally, DSM is generated at a resolution of 1m from the filtered points.

Table 2-2. LiDAR statistics

Features	Elevation	Intensity
1	Kurtosis of elevation	Kurtosis of intensity
2	Variance of elevation	Variance of intensity
3	Coefficient of variation of elevation	Coefficient of variation of intensity
4	Skewness of elevation	Skewness of intensity
5	Entropy of elevation	Entropy of intensity

Table 2-2 shows the five types of LiDAR statistics that are used in the benchmark method. These statistical features are calculated separately from the elevation and intensity of LiDAR point cloud including variance, the coefficient of variation, kurtosis, skewness, entropy, that describes elevation and intensity characteristics of the point cloud. The entropy formula is specified below:

$$Entropy = -\sum_{i=1}^{n}(P(i) * \log(P(i))) \quad (2.6)$$

where $P(i)$ is the elevation or intensity value of point i.

In the classification stage, we use a radial basis function (RBF) kernel for SVM because of its superiority of mapping feature vector in high dimensional feature space and enable better class separability (Schölkopf and Smola, 2002). During training, two parameters are optimized: the penalty parameter, C, and the kernel width, s. C provides a trade-off between the margin among classes and the number of target objects rejected. Kernel width resolves the complexity of the delineation of margins between classes (Lo and Wang, 2012). A 2D grid search method is adopted three times to find the best parameter combination. The initial searching range is a logarithmic grid from 10^{-5} to 10^{5} with 1600 combinations. After using this first round of optimization to narrow the potential value for C and s, another two rounds of searching in linear space each with 1600 combinations are used in secondary and tertiary searches to find the best parameter combination.

2.3.5 3D CNN-based multi-stage classification

The flowchart of the new method is shown in Figure 2-4. Data preparation includes data alignment of LiDAR tiles with Landsat pixels and LiDAR tiles augmentation to standardize LiDAR for feature learning. In the process of LiDAR feature extraction, voxelization is conducted to separately create the occupancy and intensity grids. The 3D CNN feature extractor

is trained based on the two types of grids. After feature extraction, extracted features are combined with spectral data for the final classification using a multi-class SVM. The same parameter tuning and accuracy assessment strategy as the benchmark method is used. The only difference is that a majority vote of sample rotations is conducted as the final step to yield the classified map. The details of major steps are described below.

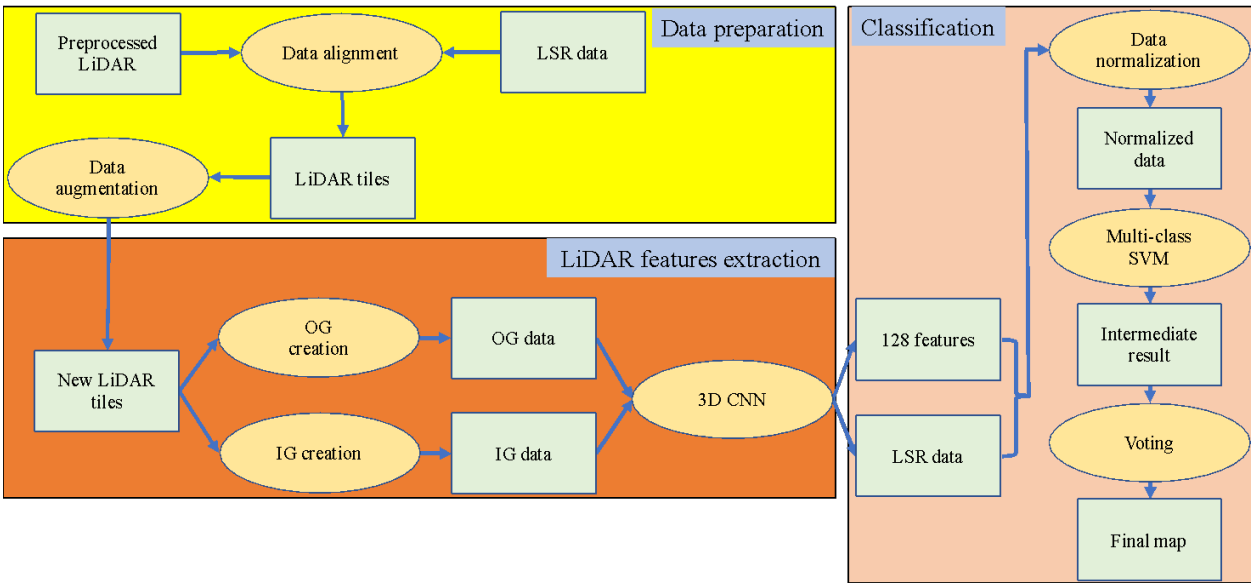


Figure 2-4. Flow chart of 3D CNN-based multi-stage classification

The same noise removal process is conducted for the new method to get rid of outlier points. After noise removal, data alignment is conducted by cropping LiDAR tiles corresponding to Landsat pixels. The point cloud cropping tool in the Pdal Library is used to extract points aligned with each Landsat pixel in the study area. In this process, boundary areas without LiDAR coverage are removed. To ensure the spatial invariant property (the ability to correctly classify samples wherever it appears in a defined space) and thus enhance the robustness of the model, data augmentation is done by rotating each training sample 9 angles around their *z*-axis. So for two neighboring samples, their rotation difference is 40 degrees (Figure 2-5). To avoid edge

point missing issues when rotating, we create 10-meter buffers for each sample in *x* and *y* direction and crop them using the sample mask after rotations.

After data augmentation, a grid size of 30×30×50 m^3 is used to standardize each tile of LiDAR at a resolution of 1 m^3. This size is chosen to ensure a balance between data quality (1 – 5 points/m^2) and computational intensity. The 1 m^3 resolution does not lead to loss of useful details since the land cover features of interest usually are larger than 1 m^3. In 3D space of each tile, *x* and *y* have the same dimension of 30 m that also equals to the size of its corresponding image pixel. The height dimension is chosen to be 50 m. Those tiles with points higher than 50 m are normalized to fit the grid. The bottom of a grid is set as the height of the lowest point within its LiDAR tile. In the model, we constructed two types of grids: occupancy and intensity grids. Figure 2-6 shows a voxelization example by constructing a grid of voxels (small unit cubes) overlaid with LiDAR points. Occupancy grid only counts the existence of LiDAR point (1 or 0) within a 1 m^3 voxel. This means no matter how many points falling within the unit, the occupancy voxel always has the value 1 and the opposite has the value 0. An intensity grid is created by calculating the average intensity values of all the points falling within the voxel, and the average value is assigned to it. If there is no point in a voxel, its value is set to 0.

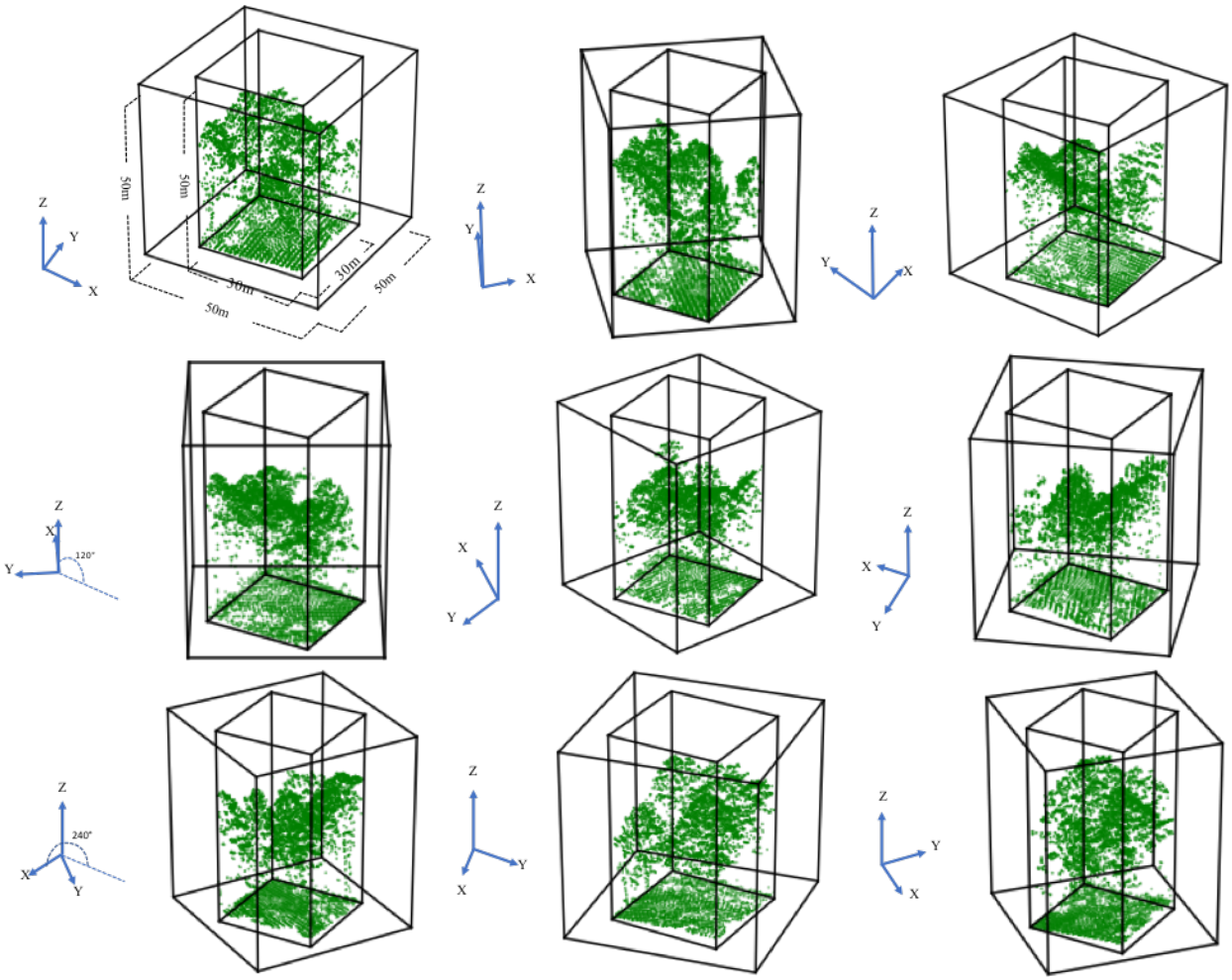


Figure 2-5. LiDAR sample rotations with 10-meter buffers in x and y (40 degrees each and totally nine directions counter clockwise)

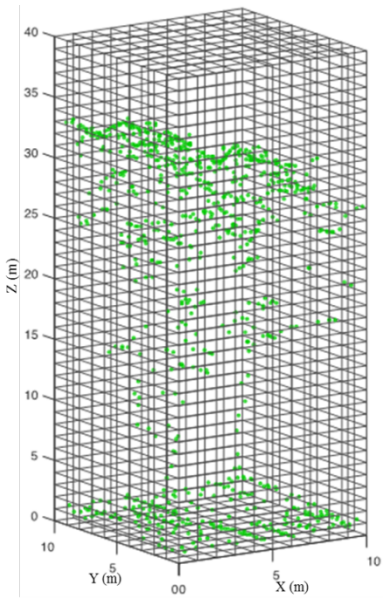


Figure 2-6. Example of the data voxelization

The structure of the 3D Convolutional Neural Network (CNN) feature extractor is shown in Figure 2-7, adapted from the Voxnet developed by Maturana and Scherer for an object classification task (Maturana and Scherer, 2015a, 2015b). The network contains two convolutional layers, one pooling layer, three dropout layers and a fully connected layer. We change this network into a feature extractor by removing the fully connected layer during the last epoch of training and extract the last dense layer as output. We also adjust the input data format and a series of model parameters in order to serve our experimentation. In Figure 2-7, the input layer accepts two fixed-size grids of x(32m)×y(32m)×z(52m) voxels and are created by adding two 1m paddings to each dimension of origin grids (30m×30m×50m). The two convolutional Layers (a, b, c) accept four-dimensional input volumes in which three of the dimensions are spatial, and the fourth contains the activation maps. These layers create c feature maps by convolving the input with c learned filters of shape d×d×d×c, where d is the spatial dimension and c is the number of input activation maps. In this network, the convolutional layer also uses a

spatial stride to reduce redundant spatial details. The first convolution layer uses 32 filters with a window size of 5×5, a stride of 2 and produce activations maps with the dimension of 14×14×24. Then the output is passed through a dropout layer with parameter 0.1. The second convolution layer accepts these and produces activation maps with the dimension of 12×12×22. The pooling Layer is used to down sample the input volume by a factor of m along the spatial dimensions by replacing each $m \times m \times m$ non-overlapping block of voxels with their maximum. In our case, the dimension of output layer from pooling is 6×6×11. Then two dropout layers are used separately with dropout rate of 0.3 and 0.4 to avoid overfitting. The fully connected layer comes the last to conduct a prediction on labels based on the combination of outputs from the previous layer (128 features).

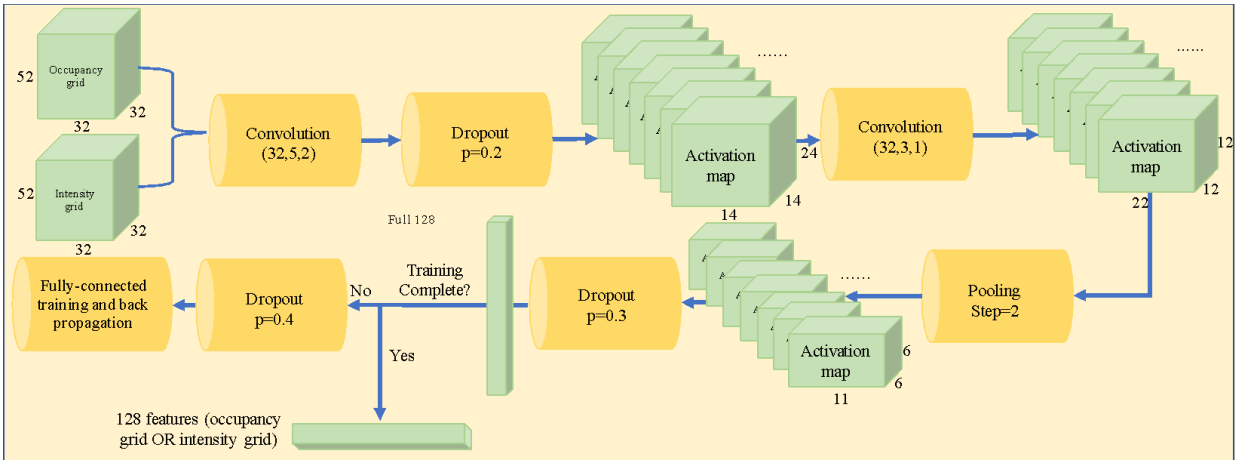


Figure 2-7. Structure of 3D CNN feature extractor

Figure 2-8 illustrates the 3D profile images of the intermediate layers generated from the network in a forest area. We can see that as the data goes through the different layers in the model, data volume reduces as tree crown features become more evident and synthesized, which is finally represented by 128 features.

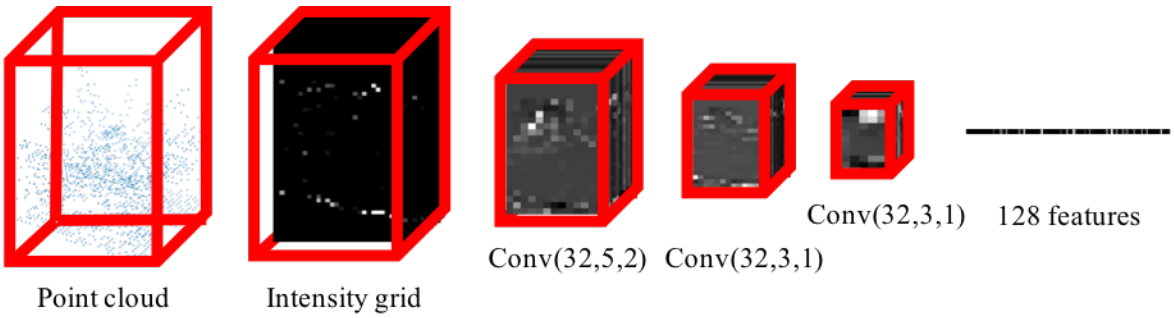


Figure 2-8. Layers in the 3D convolutional feature extractor: the input layer (point cloud of a forest sample); the intensity grid (52×32×32); the convolutional layer 1 (24×14×14); the convolutional layer 2 (22×12×12); the pooling layer (11×6×6); the layer of 128 features.

To ensure the weights of neurons in the training process are updated timely, we divide the training data into five batches within every epoch. After each training on a single batch, backpropagation is utilized to update the weights. Our model is trained 800 epochs, and the last dense layer is extracted as output instead of fitting into the fully connected layer. The reason we pick up this layer is that the 128 features are both qualitatively and quantitatively effective to aid in the classification. After feature learning, features extracted from occupancy and intensity grids are combined with spectral data for the final round of classification using SVM. Final results are achieved from a majority voting based on the votes from the nine rotation samples.

To compare the results when different numbers of images are included, four scenarios using solely LiDAR, single time image plus LiDAR, flush and senescence image pair plus LiDAR, and all images plus LiDAR are constructed. Both the benchmark method and our new model are tested, trained and evaluated using the same training, testing dataset and parameter tuning strategy with each scenario. The basic structures of these scenarios are listed in Table 2-3.

Table 2-3. Structure of classification scenarios

Table 3. Structure of classification scenarios

		Benchmark LiDAR features	IG&OG features*	Single image*	Flush and senescence image pair*	All images*
Scenario 1	F1	/				
	F2	×				
	F3		×			
Scenario 2	F1			×		
	F2	×		×		
	F3		×	×		
Scenario 3	F1			×		
	F2	×			×	
	F3		×	×		
Scenario 4	F1					×
	F2	×				×
	F3		×			×

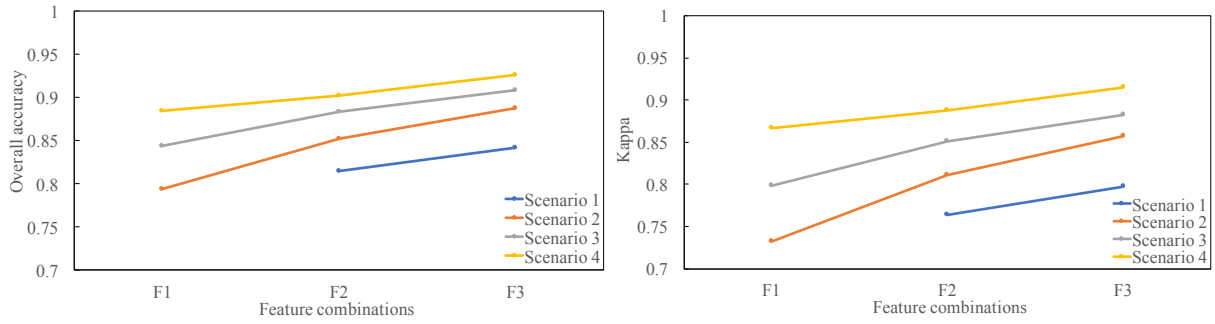
*F1-F3: feature combinations 1 to 3
*IG&OG features: 128 features extracted from both occupancy grid and intensity grid
*Single image: Spectral bands, texture statistics and NDVI calculated from Landsant 5 TM image acquired on April 17, 2011
*Flush and senescence image pair: Spectral bands, five texture statistics and NDVI calculated from two Landsant 5 TM images acquired on January 2 and June 4, 2011
*All images: Spectral bands, five texture statistics and NDVI calculated from all of the twelve Landsat 5 TM images

2.4 Results

2.4.1 Overall performance

Figure 2-9 shows overall accuracies and kappa values of the four scenarios with different feature combinations. The incorporation of CNN extracted features enhances the overall accuracy to a considerable extent. Features extracted from the new method (F3) outperform features from the benchmark (F2) method. The benchmark method performs better than using only image and image features (F1). Accuracy increase from scenario 1 to 4 also indicates that the more images incorporated, the higher overall accuracy is achieved. Specifically, the new method (F3) outperforms the benchmark method (F2) separately by 2.65%, 3.58%, 2.45%, 2.37% in overall accuracy and 0.0332, 0.0455, 0.0313, 0.0273 in kappa values from scenario 1 to 4. The benchmark method (F2) outperforms using only images (F1) from scenario 2 and 4 by the range from 2% to 6% in overall accuracy. The accuracy increases of combining benchmark features with the single image, the flush and senescence image pair, and all the images are in the

range from 2% to 4%; The increases of using IG&OG features combined with the same image combinations are in the range from 2% to 5%.

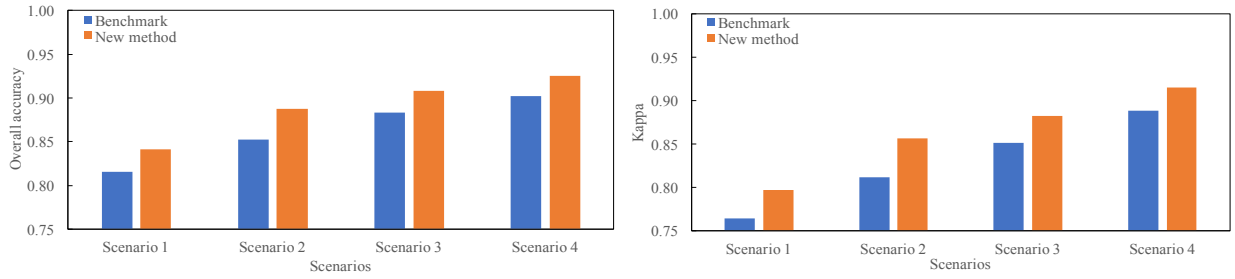


(a) (b)

Figure 2-9. The changes of overall accuracies (a) and kappa values (b) among different feature combinations

Figure 2-10 compares the new method and the benchmark method among the four scenarios. Scenario 1 compares the classification results solely generated from LiDAR features and the increase attributes to agriculture, developed, wetlands, shrub and water classes. Scenario 2 combines LiDAR features with a single image and the increase attributes to agriculture, developed, wetlands, herbaceous and shrub classes contribute. Scenario 3 uses a pair of flush and senescence images combined with LiDAR features and the increase attributes to wetlands, herbaceous, shrub and water classes. Scenario 4 combines all images and LiDAR features and the increase attributes to agriculture, developed, wetlands, herbaceous, shrub, and water classes.

In order to test the significance of the accuracy increase, McNemar's test is used and the testing result is shown in Table 2-4. The null hypothesis of the testing is that there is no significant difference between the overall accuracy of the new method and benchmark method. All of the four *p*-values are less than 0.05, which strongly rejects the null hypothesis and indicates that the increases are significant.



(a) (b)

Figure 2-10. Comparison of overall accuracies (a) and kappa values (b) between the new method (F3) and the benchmark method (F2)

Table 2-4. McNemar's test of the difference in overall accuracies between the new and benchmark method

	McNemar's chi-square	P-value
S1	64.015	1.24E-15
S2	87.011	2.20E-16
S3	59.016	1.56E-14
S4	57.017	4.32E-14

2.4.2 Classification accuracy on different classes

Figure 2-11 compares user's and producer's accuracies between the new and benchmark methods among the four scenarios. User's accuracy reflects how often the samples classified are correct and producer's accuracy reflects how often samples from the reference are correctly classified.

In general, the new method has higher accuracies than the benchmark method. In scenario 1, the highest increases in user's accuracy are developed (14.35%), wetlands (5.19%), herbaceous (4.88%), shrub (4.03%), and water (2.26%). The increases of agriculture and forest are about 1%. For producer's accuracy, agriculture, forest, herbaceous, and shrub separately increase by 7.93%, 1.28%, 0.87%, and 11.64%. Developed and wetlands decrease about 1% and

water decreases 3.24%. In scenario 2, the highest increases of user's accuracies are developed (11.09%), shrub (7.56%), agriculture (6.97%), herbaceous (4.42%), wetlands (3.29%), and forest (0.16%). Water has a decrease of 0.66%. Meanwhile, producer's accuracies have increases in agriculture (3.84%), developed (6.67%), wetlands (2.84%), herbaceous (7.78%), shrub (7.76%) and forest (0.95%). Water drops 0.65%. In scenario 3, the increases of user's accuracies from the new method are mainly developed (12.09%), wetlands (4.76%), shrub (4.76%), agriculture (3.71%). The increases of forest and water are about 1%. Herbaceous decreases (0.32%). Meanwhile, the increases of producer's accuracies are mainly in shrub (7.33%), wetlands (4.54%), developed (4.44%), herbaceous (4.04%), and forest (1.06%). The increases of water is about 1%. Agriculture remains the same. In scenario 4, user's accuracy mainly increases in shrub (7.86%), developed (6.20%), wetlands (5.74%) and herbaceous (3.19%). Agriculture and forest increase about 1% and water drops 0.55%. For producer's accuracy, the increases mainly are in developed (8.39%), shrub (7.76%), water (2.60%), agriculture (2.39%), and wetlands (1.70%). The increases of forest and herbaceous are about 1%.

Figure 2-12 shows the average classification errors of scenario 4 that has the highest overall accuracies among four scenarios. The errors of benchmark method in wetlands, shrub, developed, and herbaceous range from 10% to 24%. Agriculture, forest, and water range from 3% to 7%. Errors of the new method in agriculture, developed, wetlands, forest, herbaceous, shrub and water range from 2% to 20%. Compared to the benchmark method, the error rates of the new method separately drop by 1.23%, 7.29%, 3.72%, 0.73%, 1.88%, 7.81%, and 1.02% in agriculture, developed, wetlands, forest, herbaceous, shrub and water.

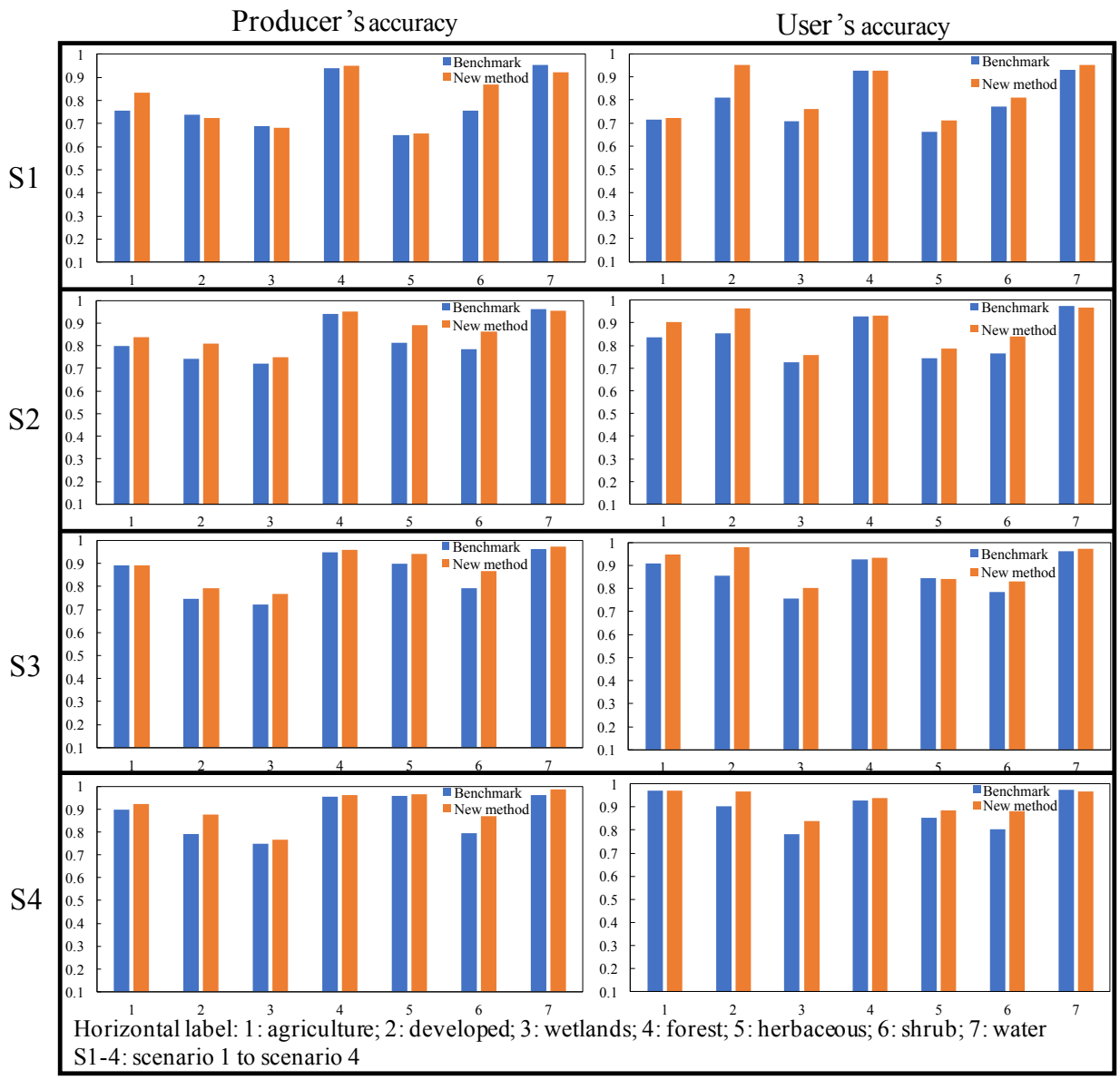


Figure 2-11. Comparison of user's and producer's accuracies between the new methods (F3) and benchmarks (F2)

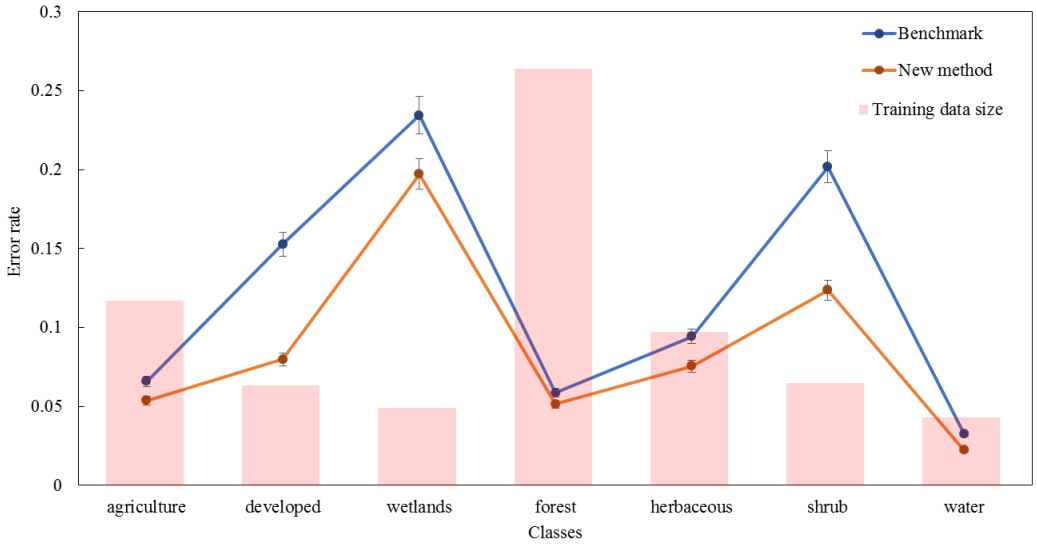


Figure 2-12. Comparison of error rates among classes between the new method (F3) and benchmark (F2) in scenario 4 (error rate: the average of commission and omission errors)

2.4.3 Characteristics of spatial distribution of classified land cover types

The final maps generated from the benchmark and new methods from scenario 1 to 4 are shown in Figure 2-14. We zoom-in four typical locations (Figure 2-13) to demonstrate our improved performance compared with the benchmark method. Areas of difference are separately extracted from the benchmark and new method and shown as colored polygons in Figure 2-13 (scenario 4) and Figure 2-15 (scenario 1). Since the four scenarios show a similar pattern, we only discuss scenario 4 (Figure 2-13) which has the highest overall accuracy. At location 1, the new method correctly extracts shrub and herbaceous in forest gaps that the benchmark wrong classifies as forest. Developed land scattered in the northeastern forest is correctly classified by the new method but is wrongly recognized as shrub or forest by the benchmark. Location 2 is an urban area. In contrast to the benchmark method, the new method can correctly classify the majority of urban pixels that the benchmark method misclassifies as herbaceous, shrub, or forest. The new method is also able to recognize urban herbaceous pixels that the benchmark method

wrongly classifies as developed class. In location 3, the new method performs better in classifying agriculture, shrub and herbaceous compared with the benchmark method that misclassifies many agriculture pixels as herbaceous and many shrub and herbaceous pixels as developed. At location 4, the benchmark method misclassifies many water pixels as developed in the center of the map. The new method can extract wetlands pixels along a river in the dense forest compared with the benchmark method which wrongly classifies them as forest.

Overall, the benchmark tends to overestimate certain classes, for example, developed and forest classes in scenario 4 and herbaceous and forest in scenario 1, by its ability of taking into account complex structural compositions of land cover. The new method can extract more complex features and thus detect reasonable patterns on classified maps.

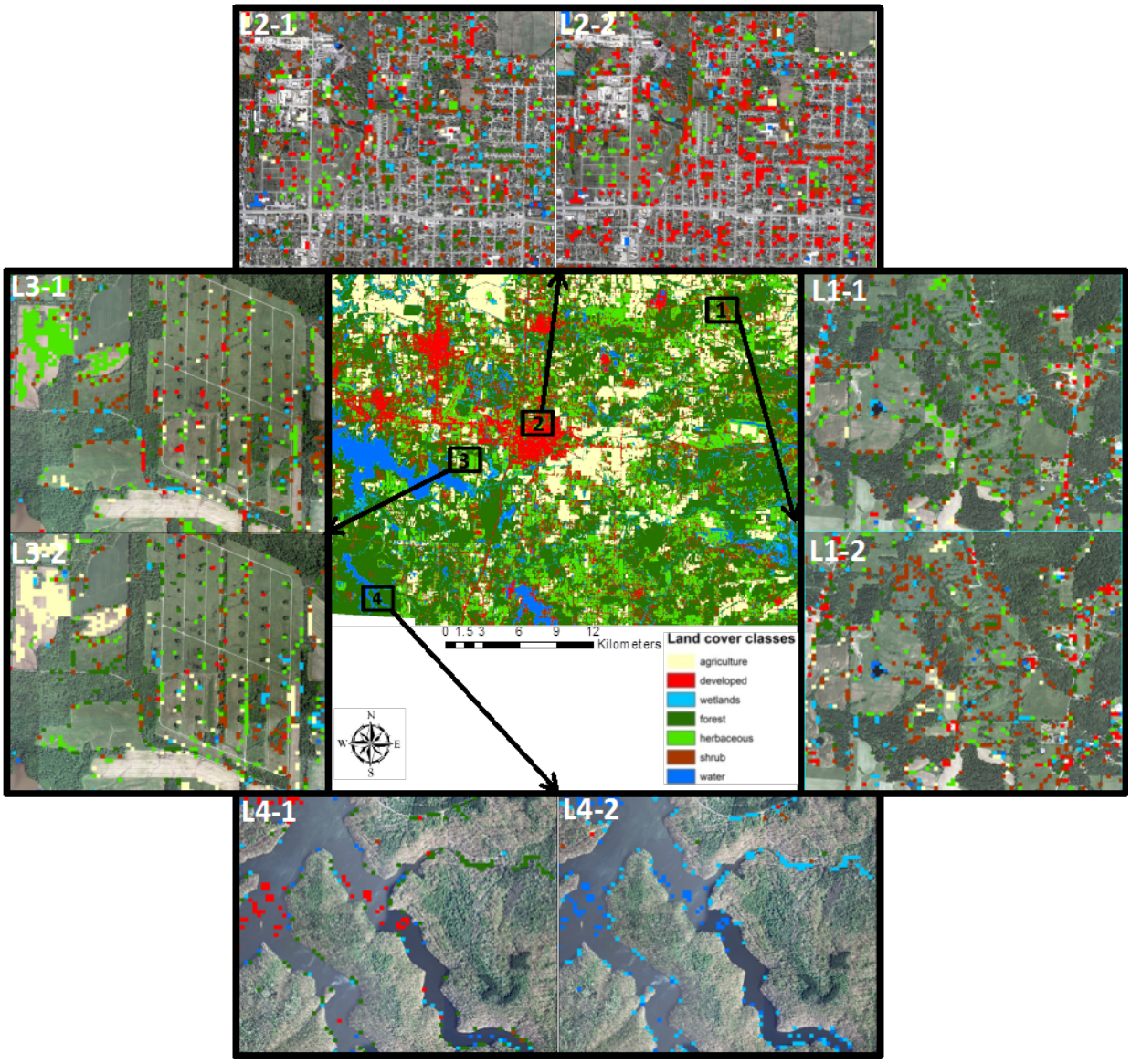


Figure 2-13. Classification difference between the best models of the benchmark method and new method (L1-1, L2-1, L3-1, L4-1: benchmark maps of location 1-4; L1-2, L2-2, L3-2, L4-2: maps from the new method of location 1-4; center map: the classified map of research area generated from the new method)

2.5 Conclusions and discussion

This paper describes a novel land cover classification method based on a 3D convolutional neural network feature extractor using comprehensive 3D geometric and intensity

information from LiDAR data, and develops a methodological framework for the fusion of extracted LiDAR features with multi-temporal images to achieve the high accuracy. We use reference data generated from both orthophotography and LiDAR for model training and testing. Specifically, four scenarios with the different feature combinations of features are constructed to evaluate the new method. In general, overall accuracies increase by 2.65%, 3.58%, 2.45%, 2.37% respectively for scenario 1, 2, 3, and 4. By testing on 2489 randomly selected samples, the best model achieves an overall accuracy of 92.57% and a kappa value of 0.9153. By solely using LiDAR derived features, our method can still achieve an overall accuracy of 84.17% with a kappa value of 0.7971, where the benchmark method can only reach 81.52% with a kappa of 0.7639.

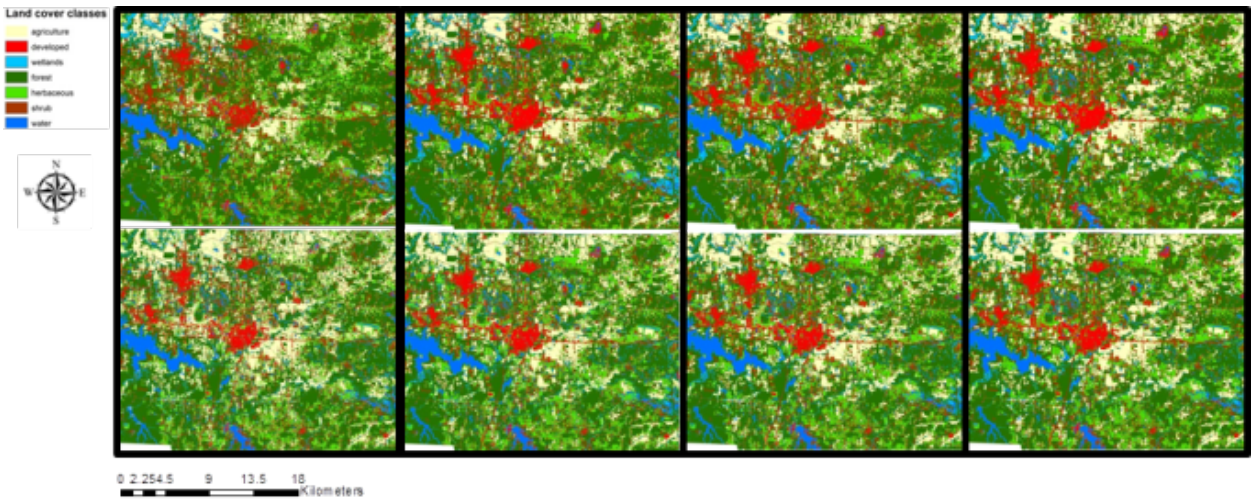
After comparing the results between the new and benchmark methods, we find that the features derived from our method perform better than benchmark method based on a 2D convolutional neural network and are promising for capturing extensive complex features of various land cover types in three dimensions that traditional methods are not designed to resolve. For example, the benchmark method fails to classify many wetlands pixels located along rivers or lakes in a dense forest, developed regions that are shaded by surrounding vegetation, shrub and herbaceous pixels that share similar spectral information and growing patterns and turbid water pixels, but they are correctly classified by our method. Our validation process using orthophotography and LiDAR further proves that the new method can achieve high-accuracy land cover mapping.

Our current implementation has the following limitations. A large training dataset is necessary in order to ensure the effectiveness of the method and errors might come from those classes with a limited number of training samples. Errors might also come from reference

imagery and LiDAR data. From the imagery perspective, limited quality of the aerial photography could cause errors. Since the airborne photography was acquired at different times of a day, the variation of flight angles could be a reason that leads to distortion of the imagery and further adds errors in visual interpretation. From the data side, LiDAR points can be missing for some places caused by occlusion from physical obstacles or large flight angles. In these cases, real features of underneath objects cannot be learned and correctly extracted, which in turn could cause errors in the results. In areas where vegetation or other coverage is dense, ground returns may also miss from detection and raise errors. Besides, misalignment between LiDAR and Landsat imagery could also cause errors.

From the landscape context perspective, our method reveals the complexity of different landscape types in 3D and provides a synthesis approach to enabling mathematical modeling of LiDAR-based voxels. Our method departs from traditional metric-oriented analysis using LiDAR and focuses on the 3D information that can be characterized by LiDAR, which enables applications at various spatial scales. Future work will focus on exploring the optimal spatial resolution of voxels, understanding the deep features extracted and testing areas with more complex landscapes.

2.6 Supplementary



(a) (b) (c) (d)

Figure 2-14. Final classified maps (a, b, c, d: separately are the benchmarks (upper) and new methods (lower) in scenarios 1, 2, 3, 4)

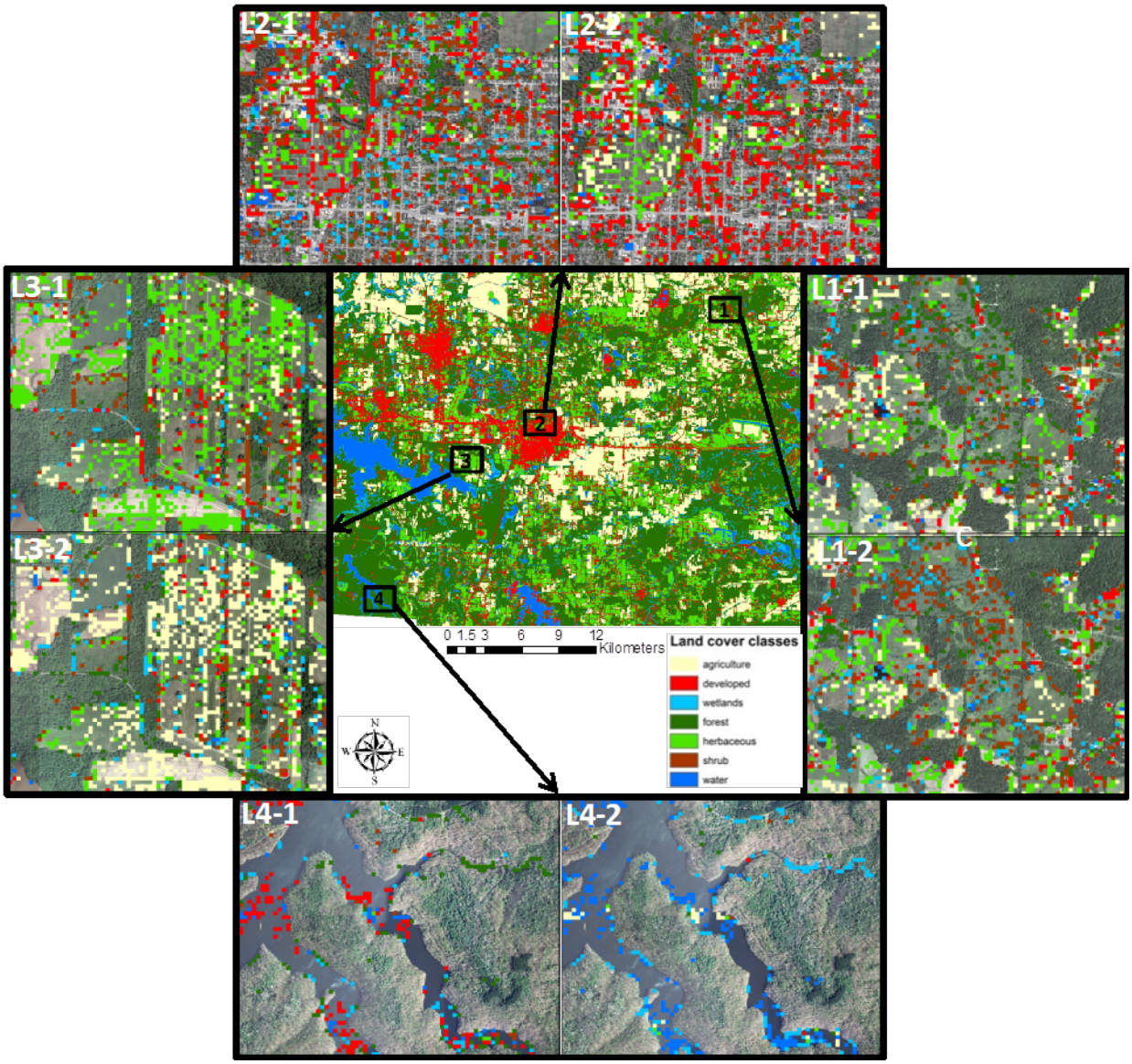


Figure 2-15. Classification difference between the best models in scenario 1 of the benchmark method and new method (L1-1, L2-1, L3-1, L4-1: benchmark maps of location 1-4; L1-2, L2-2, L3-2, L4-2: maps from the new method of location 1-4; center map: the classified map of research area generated from the new method)

Table 2-5. Accuracy assessment of benchmark scenario 1

	agriculture	developed	wetlands	forest	herbaceous	shrub	water	user's accuracy
agriculture	314	7	0	0	116	1	0	0.7169
developed	5	166	8	5	2	14	5	0.8098
wetlands	1	15	121	26	1	5	2	0.7076
forest	0	8	36	881	0	26	0	0.9264
herbaceous	94	9	0	0	225	11	0	0.6637
shrub	2	17	3	27	3	175	0	0.7709
water	0	3	8	0	0	0	147	0.9304
Producer's accuracy	0.7548	0.7378	0.6875	0.9382	0.6484	0.7543	0.9545	
Overall accuracy	0.8152							
Kappa	0.7639							

Table 2-6. Accuracy assessment of new method scenario 1

	agriculture	developed	wetlands	forest	herbaceous	shrub	water	user's accuracy
agriculture	347	11	0	0	113	3	6	0.7229
developed	1	163	2	2	0	3	0	0.9532
wetlands	0	3	120	26	2	4	3	0.7595
forest	1	12	39	893	0	15	2	0.9283
herbaceous	67	19	0	0	228	5	1	0.7125
shrub	0	16	9	18	4	202	0	0.8112
water	0	1	6	0	0	0	142	0.9530
Producer's accuracy	0.8341	0.7244	0.6818	0.9510	0.6571	0.8707	0.9221	
Overall accuracy	0.8417							
Kappa	0.7971							

Table 2-7. Accuracy assessment of benchmark scenario 4

	agriculture	developed	wetlands	forest	herbaceous	shrub	water	user's accuracy
agriculture	374	2	0	0	10	0	0	0.9689
developed	5	178	2	2	0	7	3	0.9036
wetlands	0	11	132	19	1	3	3	0.7811
forest	0	12	33	896	0	24	0	0.9285
herbaceous	35	9	0	0	333	14	0	0.8517
shrub	2	13	5	22	3	184	0	0.8035
water	0	0	4	0	0	0	148	0.9737
Producer's accuracy	0.8990	0.7911	0.7500	0.9542	0.9597	0.7931	0.9610	
Overall accuracy	0.9020							
Kappa	0.8880							

Table 2-8. Accuracy assessment of new method scenario 4

	agriculture	developed	wetlands	forest	herbaceous	shrub	water	user's accuracy
agriculture	383	3	0	0	9	0	0	0.9696
developed	0	196	0	1	1	4	1	0.9655
wetlands	0	4	135	19	1	1	1	0.8385
forest	1	9	34	902	0	17	0	0.9367
herbaceous	30	5	1	0	334	8	0	0.8836
shrub	1	7	1	17	1	202	0	0.8821
water	0	0	5	0	0	0	152	0.9682
Producer's accuracy	0.9229	0.8750	0.7670	0.9606	0.9653	0.8707	0.9870	
Overall accuracy	0.9257							
Kappa	0.9153							

CHAPTER 3: URBAN BUILDING EXTRACTION AND CHANGE DETECTION USING MULTITEMPORAL LIDAR DATA BASED ON A DEEP LEARNING AND RULE-BASED METHOD

Abstract

Urbanization has intensified across the globe at an unprecedented pace during the past several decades. Consequently, the demand for accurately acquiring urban building and building change information is expected to continue its increasing trend in the foreseeable future. Traditional methods using optical remote sensing data are limited due to image perspective variation and illumination discrepancies. The use of LiDAR data can avoid such limitations and provide fine-scale classification and change detection results. However, previous research using LiDAR was significantly restricted by its rigid requirement of registration rate, rule-based removal of non-building objects, and high sensitivity due to comparison with mixed classes of points. This research therefore develops a novel method for urban building classification and change detection through an object-based (at the level of individual buildings) approach to deriving building volume and footprint information. Two LiDAR datasets collected during 2002 and 2014 respectively in the City of Boston, Massachusetts, USA were used to evaluate the method. From the classification perspective, we established a framework by adapting a state-of-the-art point labeling algorithm (PointSIFT) to the process of building classification and achieved high accuracy for both datasets. From the perspective of change detection, our method is able to avoid typical errors and restrictions in traditional methods and achieve change detection with high accuracy in a rule-based method.

3.1 Introduction

To understand urban dynamics and urbanization often requires data-intensive change detection methods (Thakuriah et al., 2016). Urban changes affect various social, environmental, and economic conditions, including for example population migration, urban air quality, city infrastructure planning, business site locating, and urban greening. Timely characterization of urban buildings and their changes is not only beneficial to policy makers and land managers for their decision making, but also important to urban habitants for better understanding their living environments (Donihue and Lambert, 2015; Song et al., 2016; Venerandi et al., 2017). During the past several decades, remote sensing technologies have been advanced to provide an efficient and effective way for urban change monitoring. Extensive research has been conducted using optical remote sensing data to measure spectral reflectance of urban objects (Yuan et al., 2005; Bouziani et al., 2010; Resta et al., 2011; Ban and Yousif, 2012; Xu et al., 2017). The data acquired during different time periods often needs to be corrected for radiation differences caused by different sun angles, weather, and sensor conditions (Deng et al., 2017).

However, optical remote sensing data are not straightforward to capture urban information in the vertical direction, which represents an important part of urban changes taking place vertically (e.g., building construction and demolition, infrastructure construction, vegetation change). In contrast to the optical data, airborne LiDAR sensors can obtain 3D information in urban environments with fine accuracy and details (Yan et al., 2015). Consequently, in recent years, airborne LiDAR technology has gained rapid progress and enabled many urban applications (Zelener and Stamos et al., 2016; Alonzo et al., 2015; Hata and Wolf, 2016). Thus, such LiDAR data provide a new avenue for urban change detection.

LiDAR data consist of irregularly-distributed points which contain not only 3D coordinates (x, y ,z), but also a series of attributes including reflectance intensity, incidental angle, GPS time stamp, scan angle, number of returns, etc. Since LiDAR points are often dense, and irregularly distributed in space, it is difficult to use traditional spatial data structures for supporting analytics (Barber et al. 2008; Richter et al. 2013; Xu et al., 2015). This challenge calls for cutting-edge research to fully exploit the advantages of LiDAR (Qi et al., 2017a). A straightforward strategy is to transfer points to rasters and conduct change detection analysis by calculating the difference between rasters (Teo and Shih, 2013). Another way is to construct voxel grids to synthesize irregularly distributed points and detect changes by checking the status of existence of points within voxels since changed voxels would have a status change during multi-temporal periods (Xu et al., 2015). However, the relationship among points is often implicit and difficult to be fully represented using traditional 2D features (Du et al., 2016) and voxel-based features (Papon et al., 2013; Maturana and Scherer, 2015a; Xu et al., 2018). A recent study (Xiao et al., 2015) combines an occupancy grid and distance-based method for a general-level 3D change detection using terrestrial LiDAR point clouds and achieve good accuracy. However, this method fails to distinguish static changes like buildings from temporary changes including moving cars, pedestrian, bikers, etc. Furthermore, since the accurate calculation of distance and angles are required, this method is not resilient to noises of measurements and misalignment of multitemporal datasets.

In order to harness point cloud information to the best degree possible, several point-based deep learning models have been developed. One pioneer approach is called PointNet (Qi et al., 2017a), which takes raw point cloud as input instead of voxels or mesh to extract features directly. It not only accelerates computation but also notably improves the accuracy of many

point-based classification tasks (Qi et al., 2017a; Yousefhussien et al., 2018). PointNet++ is an advanced version of PointNet, which incorporates a hierarchical structure of point neighborhood learning through downsampling and interpolation (Qi et al., 2017a). However, both PointNet and PointNet++ are limited in their ability to capture complex shape patterns based on their simple design of orientation-encoding (*k*-nearest neighbour searching) and weak scale awareness for feature calculation (Jiang et al., 2018).

The scale-invariant feature transform (SIFT) excels over many other feature encoding methods by using the strategy of multi-orientation feature encoding and a scale-aware design (Furuya and Ohbuchi, 2009; Darom and Keller, 2012). PointSIFT, which applies SIFT to 3D point clouds, has shown its robustness by outperforming the state-of-the-art methods including PointNet and PointNet++ on S3DIS (Armeni et al., 2016) and ScanNet (Cohen et al., 2018) datasets (Jiang et al., 2018). PointSIFT architecture incorporates 8-direction orientation-encoding (OE) units from multiple spatial scales into deep neural networks. In this way, the neurons in the stacked OE units can perceive different scales. These OE units are closely connected in the neural network and the appropriate scale would be automatically selected during model training (Jiang et al., 2018). Like PointNet++, PointSIFT also utilizes the two-stage (i.e., encoding (downsampling) and decoding (upsampling)) for feature learning. It significantly improves the feature extraction ability of the network (Jiang et al., 2018).

In this research, we develop a deep learning method to do a point-based urban building classification by adapting the PointSIFT algorithm to airborne LiDAR. Then we use an Euclidean-distance based clustering method to separate individual buildings and estimate building footprints. Finally, we infer building changes and classify change types by estimating the change of building volume and footprint from 3D alpha shapes, which is a generalization of

the 3D convex hull and a subgraph of the Delaunay triangulation over a set of 3D points. Four types of changes including demolition, new construction, reduction, and expansion are extracted based on their different patterns. A reference dataset which is created by visual interpretation from a multitemporal aerial photography and LiDAR data are used to separately evaluate the recall (completeness), precision (correctness), and the F1 score of building classification and change detection results. Two LiDAR datasets collected during 2002 and 2014 respectively in the City of Boston, Massachusetts, USA were used to evaluate the method. The estimated volumes in 2014 are also validated from a 3D building surveying data from the government data portal of Boston. The central contributions of this research are summarized as follows.

1. This research adapts the cutting-edge PointSIFT deep learning algorithm (PointSIFT) for point labeling to airborne LiDAR classification by integrating the iterative farthest point algorithm for training sample standardization, a multi-patch-based classification strategy, and a point-based voting based on different sizes of training patches to achieve high accuracy of classification.

2. Our research conducts building change detection by estimating the change of building volumes and footprints. It provides a new way for quantitatively estimating urban changes in 3D at a fine spatial scale.

3. Our method eliminates the computation-intensive point cloud registration in previous studies, by directly working on individual buildings, where the footprints of individual buildings are delineated and overlapping ones are formed into pairs for comparison of changes. This design avoids the difficulty of distinguishing between non-building and building changes, which many previous studies struggled with.

4. Although high quality (+20 pts/m^2, <0.1 RMSE) multitemporal LiDAR data have been proven to be effective for 3D urban modeling and change detection (Qin et al., 2016), this research demonstrates that it is still possible to get high accuracy results by using readily available LiDAR data (3 pts/m^2) with a large error rate (0.5 RMSE).

3.2 Background

3.2.1 Study area

Our study area covers 9 km^2 and contains the central part of eastern Boston downtown and part of southern Boston consisting of industrial and residential areas. This area has an elevation ranging from 0 to 32 m above sea level. This area is ideal to perform this research due to significant land cover changes from 2002 to 2014 with available multi-temporal LiDAR data. This area also has a complex urban building composition and there is a pressing need for improved characterization of urban dynamics to accurately measure the locations and amount of changes (Glaeser and Ward, 2009). There were multiple large construction projects carried out in this area from 2002 to 2014, including the Central Artery/Tunnel Project (the "big dip": https://en.wikipedia.org/wiki/Big_Dig) and the Boston Convention and Exhibition Center Project (https://en.wikipedia.org/wiki/Boston_Convention_and_Exhibition_Center).

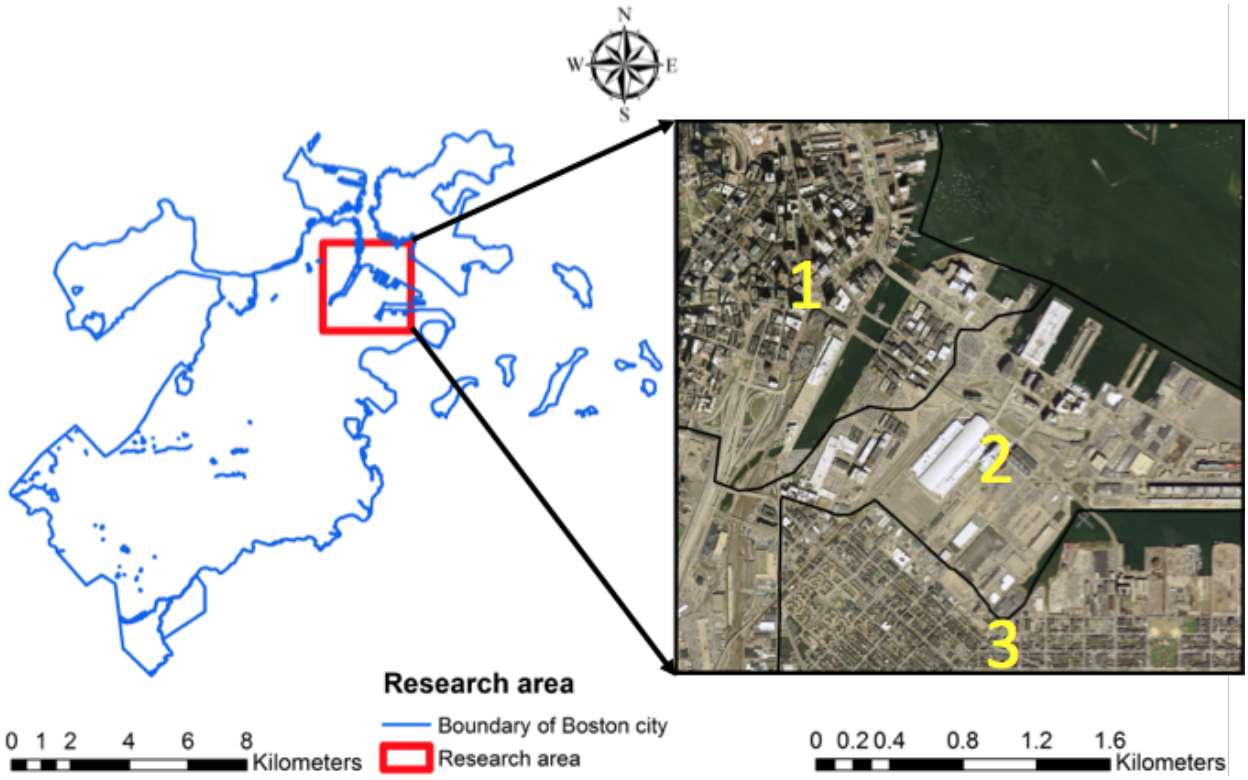


Figure 3-1. Study area (The orthophotography is collected by USGS in 2013/2014: 1: commercial area; 2: industrial area; 3: residential area).

3.2.2 Multi-temporal LiDAR dataset

There are two separate LiDAR acquisitions (2002 and 2014). The 2002 LiDAR data are provided by the Massachusetts Executive Office of Technology Services and Security (EOTSS:https://www.mass.gov/orgs/executive-office-of-technology-services-and-security). They were collected by 3Di Technologies, Inc. (Now Spectrum Mapping LLC) using 3Di's Digital Airborne Topographic Imaging System II (DATIS II) at the height of 2000m in the date range between April 7th and June 25th in 2002. The point density is 3 pts/m^2 with a horizontal RMSE of 0.5m and vertical RMSE of 0.15m. The 2014 LiDAR data are from the NOAA Digital Coast Data Access Viewer (https://coast.noaa.gov/dataviewer/#/lidar/search/). They are from the New England CMGP Lidar Processing Project for the United States Geological Survey (USGS).

The 2014 LiDAR data were collected using a Leica ALS70 sensor and an Optech ALTM Gemini LiDAR sensor. Both sensors collect up to four returns (echo) per pulse, as well as intensity data at an average height of 1753m on April 7th in 2014. The point density is 3pts/m^2 with a horizontal RMSE of 0.21m and vertical RMSE of 0.024m.

3.2.3 Reference dataset

Four areas (red squares in Figure 3-2) are selected as reference locations for model training. The 3D LiDAR points in these four different areas are manually labelled by using ArcGIS Las dataset tool box (profile and 3D visualization functions), where building points are labeled as 1 and non-building points as 0. Considering the data quality variation between 2002 and 2014 datasets, training data were generated separately for 2002 and 2014.

For the purpose of validating detected building and changes, two sets of orthophotography are downloaded from the MassGIS website using the OLIVER: MassGIS's Online Mapping Tool. The 2001 orthophotography was collected by the MassGIS and the Massachusetts Department of Transportation and includes a set of true color imagery (composed of red, green, and blue band) with a resolution of 0.5m. The 2013/2014 orthophotography was collected by USGS and contains a set of color-infrared (CIR) imagery (composed of red, green, blue, and near-infrared (NIR) band) with a resolution of 0.3m. A cross-checking process is conducted to delineate changed building locations. These locations are further verified based on the multitemporal orthophotography from Google Earth Engine Historical Imagery Archive (https://earthengine.google.com/) (Gorelick et al., 2017).

Figure 3-2. Training data for building extraction (1: downtown area; 2: coastal area;3: warehouse area; 4: residential area).

3.3 Methods

3.3.1 Data preprocessing

The 2002 LiDAR dataset was provided with the original format of .csv with *x*, *y*, *z* coordinates, intensity, and return number as the attributes. The 2014 LiDAR dataset was provided with the standard format of .las with the same attributes. Since the intensity information of the 2002 dataset comes in a raster format, a bilinear interpolation is used to interpolate intensity values for each point based their x and y coordinates. Return number from the 2002 LiDAR is disrupted so we did not include it in the classification. Apart from this, we also calculate the local surface curvature and coefficient variation of each point with its surrounding points. Firstly, we estimate covariance matrix *C* considering the point and its neighboring points within a distance of 20m, which allows the points to form a reasonable surface structure. Then, we calculate the three eigenvalues from the covariance as λ_0, λ_1, λ_2. Finally, the curvature value

is calculated based on the eigenvector calculated. The formula for calculating covariance and curvature is listed below:

$$Covariance = \frac{1}{n}\sum_{i=1}^{n}\left(P_i - \vec{P}\right)\left(P_i - \vec{P}\right)^T,\ C \cdot \vec{V}_j = \lambda_j \cdot \vec{V}_j,\ j \in \{0,1,2\} \quad (3.1)$$

$$Curvature = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \quad (3.2)$$

where n is the number of point neighbors considered in the neighborhood of P_i. $\vec{P}$ represents the 3D centroid of the nearest neighbors within the distance of 20m, λ_j is the j-th eigenvalue of the covariance matrix, and $\vec{V}_j$ is the j-th eigenvector. λ_0, λ_1, λ_2 are the three eigenvalues of the covariance matrix.

Coefficient of variance (CV) is calculated from the angle difference of the eigenvector between the center point and its neighbors. Six bins (0°-30°, 30°-60°, 60°-90°, 90°-120°,120°-150°, 150°-180°) are created to count the frequency of neighboring points falling in each of the bins. CV is then estimated from the variance and mean values.

After transferring LiDAR into standard .las format, Lastools (Isenburg, 2011) is used for ground/non-ground classification and height normalization. A standardized processing chain for filtering ground points was developed for the study area. Raw LiDAR points were first classified into ground points and non-ground points, and then height-normalized by subtracting the ground heights from the orthometric heights at the corresponding locations. The normalized LiDAR points with above-ground heights were then noise-filtered by removing outlier points with suggested parameters from the PDAL library (https://pdal.io/).

3.3.2 Point-based classification

The point-based classification aims to label all the points in 3D space. In this paper, we utilize PointSIFT (Jiang et al., 2018), which is shown in the middle part of Figure 3-4 as our classifier for extraction of building points. PointSIFT is a local feature description method that models various orientations and is invariant to scale. In the model, the PointSIFT module is used to extract abstract point features among a series of set abstraction (SA) and feature propagation (FP) modules. The set abstraction (SA) modules are used to shrink the size of the point set to 1024, 256, 64 respectively. Then the feature propagation (FP) modules are used to upsample points to the size of 256, 1024, and 8192 respectively. Finally, point features from the last FP module are fed to the fully connected layer for label prediction of points (Jiang et al., 2018). Unlike indoor scanning dataset including S3DIS (Armeni et al., 2016) and ScanNet (Cohen et al., 2018), remote sensing data are collected continuously over large geographic areas. Thus, PointSIFT is not directly applicable to the data. We split the continuously collected data into different sizes (20m, 25m, 30m) of tiles at a certain buffer size (4m, 5m, 6m) because this range covers the majority of building sizes in our study area after testing. Since each tile needs to have the same number of points as inputs, we utilize the iterative farthest point sampling (IFPS) method to resample each tile to the average point number among tiles (Moenning and Dodgson, 2003). Compared with random sampling, IFPS is able to identify those points that best preserve shape features of ground objects. The average point numbers for the patch size of 20m, 25m, and 30m are 1468, 1929, 2391 for 2002 and 1538, 2403, 3461 for 2014. The pseudo code for the IFPS algorithm is provided as follows.

1. Create an empty list of selected points S and append a randomly selected point from the total point set R, at the same time, remove it from the total point set; initialize the count number C as 1 to indicate when to stop the loop as C reaches the desired point number K.
2. Calculate the distances from the picked point in S to every point $P_i \in T$ and store them in a list called D.
3. For every point $P_i \in T$:
    a. For every point $p_j \in S$:
        i. Calculate the distance between p_j and P_i, assign it to a variable called m, and update the D_i with the minimum value between D_i and m.
4. Remove the point with the index that has the maximum value in D from R and append the point to S.
5. C equals $C+1$.
6. If C equals K:
    a. End loop and return S;

    else:
    b. Go back to step 2.

For model training, we get 1343, 892, 612 patches for 2002 and 1348, 896, 616 patches for 2014. The data are evenly split into five portions and four of them are used as training and one for validation. Python 3.5, Cuda 8.0, and Tensorflow 1.4.0 are used in a Linux environment for running the experiment. Each epoch of training takes about 1 minutes and it takes 12 hours in total.

3.3.3 Euclidean clustering and 3D building model construction

An Euclidean clustering method from the Point Cloud Library (http://pointclouds.org/) is used to isolate individual buildings or building groups (Rusu et al., 2011). In a densely distributed residential area, buildings or houses that have a distance of smaller than 3m are grouped together and treated as the same building unit. The clustering is conducted over 3D space and then the x and y coordinates of each cluster are used to generate individual building footprints by calculating their 2D concave hulls.. The building footprints from 2002 and 2014 are then separately merged and saved as two polygon shapefiles for further comparison. Overlapping footprints are then dissolved into one with island polygons removed. Then, building units from the two time periods are formed into overlapping pairs for further detection of area and volume changes. Finally, 889 pairs are formed by comparing 2002 and 2014 datasets.

Within each group of pairs, 3D alpha shapes are formed for each of the building units and volumes are estimated for each of them. Due to the limited point density of both LiDAR datasets, most facade points of buildings are missing, which would cause an inaccurate construction of the 3D alpha shapes of buildings. We develop a method to compensate this artefact, which is shown in Figure 3-3. We first project all the points in 3D to a 2D plane with relative elevation of 0 (ground), then we create artificial points from each ground point to its highest building points with a step size of 1m. Finally, the 3D alpha shape is constructed based on the compensated points and the volume is calculated based on the 3D alpha shape.

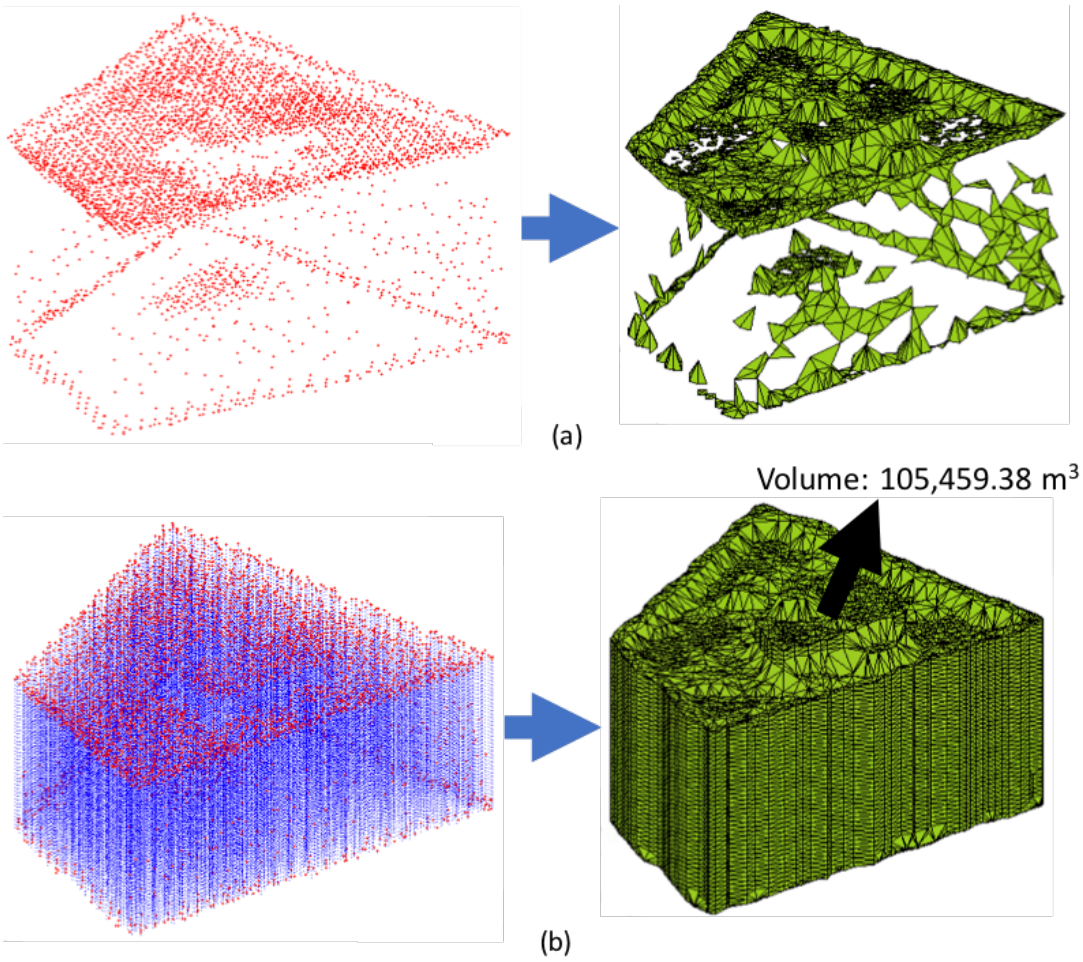


Figure 3-3. 3D alpha shape model of building (a: model generated from original points; b: model generated from vertically compensated points).

3.3.4 Building change detection and 3D change volume estimation

After deriving the 3D alpha shape for every pair of building groups, the volumes are calculated from the 3D alpha shapes. Then, the calculated values from 2002 and 2014 are compared with each other. Since we do not have a volume estimation reference for 2002 LiDAR and the accuracy of 2002 LiDAR is relatively lower, we calculate a volume normalization factor to normalize 2002 data based on 18 randomly picked buildings (6 from each of the three areas) that have no change between 2002 and 2014. The calculated normalization factor is 0.91 and the

2002 building volumes are normalized by multiplying with this factor before compared to the 2014 volumes.

After evaluating ten randomly selected pairs of changed buildings over the reference data, a 20% threshold for the volume difference and 30% for the footprint area mismatch are used as the standard for identifying changed buildings to avoid over-estimation caused by small and temporary changes of buildings (temporary goods or decorations). If any building pair has more than 20% change in volume or 30% change in its footprint area, the pair is defined as change. Different change types are classified based on the following standard:

1. If the building volume or footprint area increases from 2002 to 2014, the change type is classified as “expansion”.
2. If the building volume or footprint area decreases from 2002 to 2014, the change type is classified as “reduction”.
3. If the volume of 2002 building is 0, the change type is classified as “new construction”.
4. If the volume of 2014 building is 0, the change type is classified as “demolition”.

The flowchart of all the steps is shown in Figure 3-4 below:

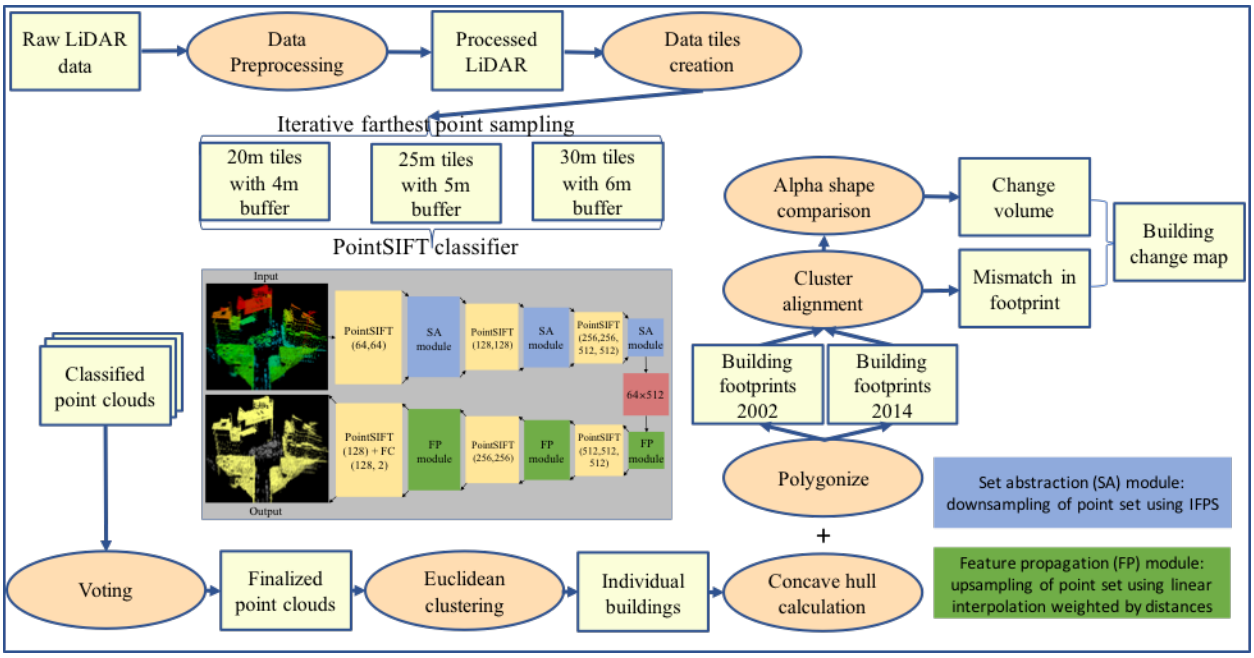


Figure 3-4. Workflow of the new method for building classification and change detection.

3.3.5 Validation

Reference data for 2002 and 2014 are visually interpreted and separately delineated through visual interpretation according to the multitemporal orthophotography and point cloud visualization tool in ArcGIS. Then, changed buildings are specifically labeled by comparing the orthophotography and LiDAR datasets from the two time periods. In this research, we define building as any human construction (except high-rise bridge) that has a footprint area of larger than 20m^2. High-rise bridge and railway construction areas are specifically filtered from the datasets through visual interpretation.

3.4 Results

Through visual interpretation, there are totally 950 buildings delineated from 2002, 1010 buildings from 2014, and 334 changed building locations found as the reference data.

3.4.1 Accuracy evaluation of building classification

Through visual interpretation, there are totally 950 buildings delineated from 2002, 1010 buildings from 2014, and 334 changed building locations found as the reference data. Figure 3-5 shows building points extracted by the classifier in both 2002 (green) and 2014 (blue). A zoom-in window is used to show the details of the classified points from nadir and 3D views, where building points are labeled with yellow color and non-building points with gray color. Figure 3-6 shows an example of the training and validation accuracies with epochs. We can see that the two accuracies converge and saturated at around the 800th epoch.

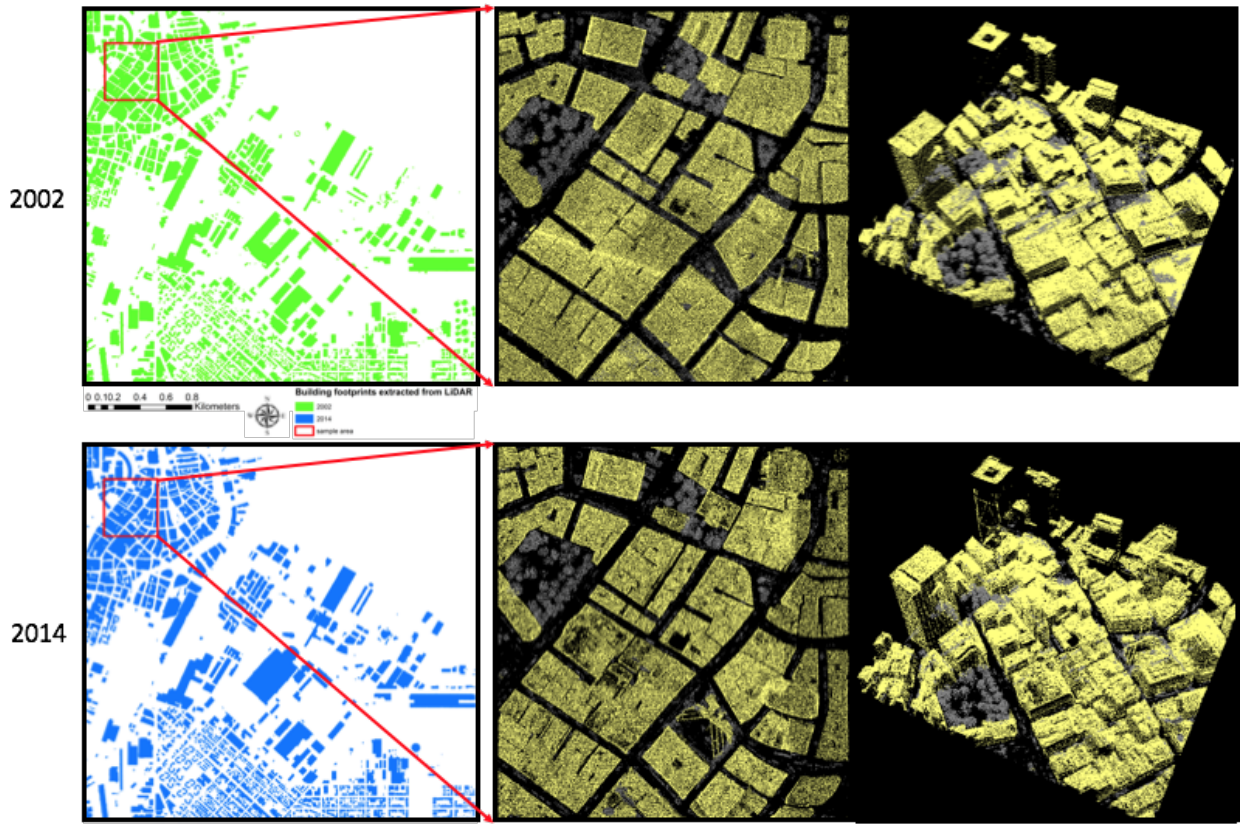


Figure 3-5. Extracted building points using patch-based PointSIFT method.

Accuracy assessment is conducted in two ways: one is for all the buildings with the area of footprint larger than 20m^2; and the other one is for buildings that have the area of footprint larger than 100m^2. The accuracy assessment of building extraction results is separately shown in

Table 3-1 and 3-2. There are 950 buildings with the area of footprint larger than 20m^2 and 810 buildings larger than 100m^2 in 2002; and our method successfully extracts 872 of them in the former group and 767 in the latter group (Table 3-1). In the group of buildings with the area of footprint larger than 20m^2, the overall recalls (completeness), precision (correctness), and F1-score are 91.79%, 98.64%, and 95.09% respectively. In terms of accuracy of the three classes of buildings, commercial building has the highest recall value (96.47%), followed by industrial building (90.86%), and residential building (89.81%).

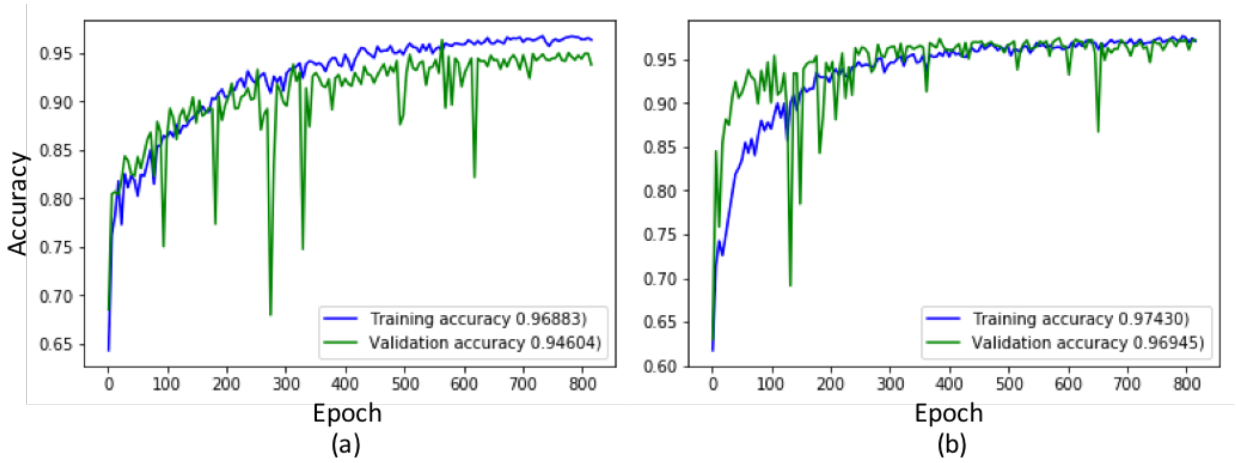


Figure 3-6. Training and validation accuracies of patch size 20 (a: 2002; b: 2014).

Table 3-1. Accuracy assessment of building extraction from the 2002 LiDAR dataset.

Area (m^2)	Type	Commercial building	Industrial building	Residential building	Overall
>20m^2	Reference	255	175	520	950
	Correct detection	246	159	467	872
	Wrong detection	5	0	7	12
	Missing detection	9	16	53	78
	Precision	98.01%	100.00%	98.52%	98.64%
	Recall	96.47%	90.86%	89.81%	91.79%
	F1 score	97.23%	95.21%	93.96%	95.09%
>100m^2	Reference	241	135	434	810
	Correct detection	233	129	405	767
	Wrong detection	2	0	1	3
	Missing detection	8	6	29	43
	Precision	99.15%	100.00%	99.75%	99.61%
	Recall	96.68%	95.56%	93.32%	94.69%
	F1 score	97.90%	97.73%	96.43%	97.09%

Table 3-2. Accuracy assessment of building extraction from the 2014 LiDAR dataset.

Area (m^2)	Type	Commercial building	Industrial building	Residential building	Overall
>20m^2	Reference	266	179	565	1010
	Correct detection	254	161	540	955
	Wrong detection	2	1	3	6
	Missing detection	12	18	25	55
	Precision	99.22%	99.38%	99.45%	99.38%
	Recall	95.49%	89.94%	95.58%	94.55%
	F1 score	97.32%	94.43%	97.47%	96.91%
>100m^2	Reference	245	147	490	882
	Correct detection	243	144	485	872
	Wrong detection	2	1	0	3
	Missing detection	2	3	5	10
	Precision	99.18%	99.31%	100.00%	99.66%
	Recall	99.18%	97.96%	98.98%	98.87%
	F1 score	99.18%	98.63%	99.49%	99.26%

Regarding precision, industrial building is the highest (100%), followed by residential building (98.52%), and commercial building (98.01%). Commercial building has the highest F1 score (97.23%), then industrial building (95.21%) and residential building (93.96%). In contrary, the group of buildings with the area of footprint larger than 100m^2 generally provides higher accuracies. To be specific, the increase of overall recall, precision, and F1-score are 2.9%,

0.97%, and 2.0% respectively. In terms of different building classes, the increases of recall are 0.21%, 4.7%, and 3.51% for commercial, industrial, and residential buildings. The increases of precision are 1.14%, 0.0%, and 1.23% for commercial, industrial, and residential buildings. The increases of F1 score are 0.67%, 2.52%, and 2.46% for commercial, industrial, and residential buildings.

A similar trend is also observed in the case of the 2014 dataset. There are 1010 buildings with the area of footprint larger than $20m^2$ and 882 buildings larger than $100m^2$; and the method successfully extracts 955 out of them in the former group and 872 in the latter group (Table 3-2). In the group of buildings with the area of footprint larger than $20m^2$, the overall recalls (completeness), precision (correctness), and F1-score are 94.55%, 99.38%, and 96.91% respectively. In terms of accuracy of the three classes of buildings, residential has the highest recall value (95.58%), followed by commercial (95.49%), and industrial (89.94%). As for precision, residential is the highest (99.45%), followed by industrial (99.38%) and commercial (99.22%). Residential has the highest F1 score (97.47%), then commercial (97.32%) and industrial (94.43%). In contrary, the group of buildings with the area of footprint larger than $100m^2$ generally achieves higher accuracies. Specifically, the increase of overall recall, precision, and F1-score are 4.31%, 0.28%, and 2.35% respectively. In terms of different building classes, the increases of recall are 3.69%, 8.02%, and 3.40% for commercial, industrial, and residential buildings respectively. The precisions of commercial and industrial buildings drop 0.04% and 0.07% respectively while residential buildings increase 0.55%. The increases of F1 score are 1.87%, 4.2%, and 2.01% for commercial, industrial, and residential buildings respectively.

3.4.2 Accuracy assessment of building volume estimation

The reference data for volume estimation were downloaded from the website of the Boston Planning and Development Agency (http://www.bostonplans.org/). The data are in the format of 3D shapefile and were surveyed in the years of 2012 and 2017. Considering the time mismatch between the reference year and the LiDAR data, we randomly chose 10 samples from each of the building types that have no change through the years of 2012 to 2017. The volume of each sample of buildings is calculated using ArcGIS 3D analyst tool. The estimated volumes calculated from 3D alpha shapes are then compared with the reference volume for accuracy assessment. The results are shown in Table 3-3. Among the three different classes, industrial buildings achieve the lowest error rate (3.24%), followed by commercial building (3.47%), while residential has the largest error rate (12.45%).

Table 3-3. Error rate of volume estimation

Sample number	Commercial building	Industrial building	Residential building
1	0.43%	3.14%	4.26%
2	8.08%	5.36%	28.36%
3	6.80%	-0.26%	8.15%
4	5.03%	-2.41%	11.58%
5	-0.65%	-4.57%	-10.14%
6	-0.62%	4.40%	16.23%
7	3.21%	6.88%	-11.96%
8	0.33%	-1.36%	-12.56%
9	4.84%	3.45%	-7.04%
10	4.68%	0.55%	14.24%
Average error rate	3.47%	3.24%	12.45%

3.4.3 Accuracy evaluation of building change detection

Changed buildings are extracted based on the change of volumes and footprint mismatches. Figure 3-7 shows the distribution of different types of building changes including demolition, new construction, expansion, and reduction. According to Table 3-4, the changes in

the commercial area are mostly expansion (21) and new construction (20). There are also some cases of demolition (11) and reduction (8). In industrial areas, the number of demolition (34) and new construction (29) is most significant. Expansion is 17 and reduction is 7. In the residential area, there are many expansion cases (63), a decent number of reduction (40) and demolition (35) cases while new construction (30) comes the last.

Table 3-4. Distribution of different types of changes (area of building footprint > $20m^2$).

Type	Expasion	Reduction	New construction	Demolition
Commercial building	21	8	20	11
Industrial building	15	7	29	34
Residential building	63	40	30	35

Table 3-5 shows the number of changed buildings among the three types of buildings. There are separately 313 and 262 detected changes of buildings with the areas of footprint larger than $20m^2$ and $100m^2$. In the group of buildings with the area of footprint larger than $20m^2$, the largest number of changes is building expansion, which is 99. Building demolition comes the second with the number of 80. New construction has a number of 79 and reduction is the lowest with the number of 55. In the group of buildings with the area of footprint larger than $100m^2$, the largest number of changes is also building expansion, which has the number of 98. New construction comes the second with the number of 61. Building reduction has the number of 52 and demolition is the lowest with the number of 51.

Table 3-5. Distribution of changed buildings.

Type of change		Expasion	Reduction	New construction	Demolition	Total
Count	$>20m^2$	99	55	79	80	313
	$>100m^2$	98	52	61	51	262

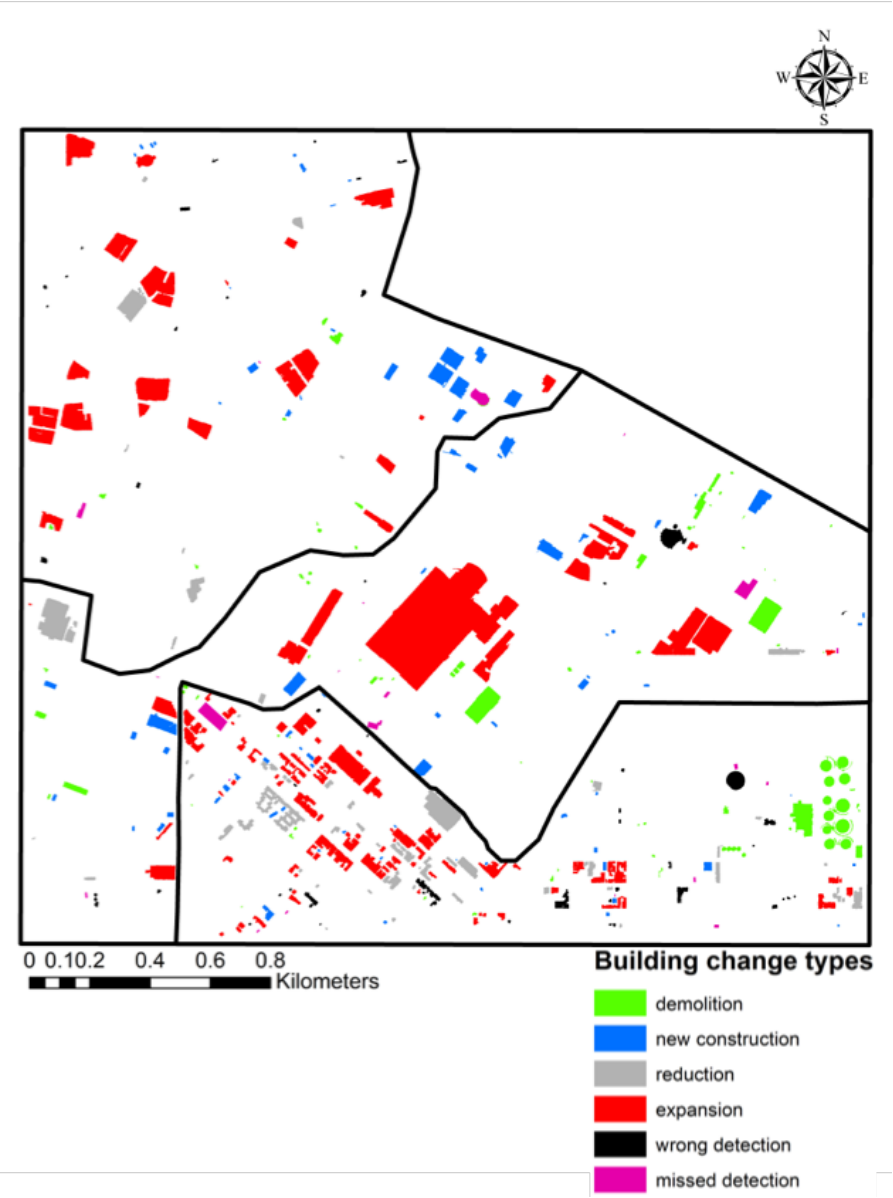


Figure 3-7. Distribution of different types of building changes (1: commercial area; 2: industrial area; 3: residential area).

According to Table 3-6, in the group of changed buildings with the area of footprint larger than 20m^2, the overall recall (completeness), precision (correctness), and F1-score are 93.71%, 86.94%, and 90.20%, respectively. In terms of the change detection accuracy for the three types of buildings, commercial building has the highest recall value (95.24%), followed by residential building (94.92%), and industrial building (90.43%). Regarding precision, industrial

building is the highest (92.39%), followed by commercial building (81.08%), and residential building (86.60%). Industrial building has the highest F1 score (91.40%), followed by residential building (90.57%) and commercial building (87.59%). For the group of buildings with the area of footprint larger than 100m^2, the method achieves higher change detection accuracies. Specifically, the increase of overall recall, precision, and F1-score are 2.26%, 4.99%, and 3.71% respectively. In terms of different building types, the increases of recall are 1.13%, 2.43%, and 2.38% for commercial, industrial, and residential buildings. The increases of precision and F1 score are separately 5.80%, 1.81%, 6.31%, and 3.79%, 2.13%, 4.48% for commercial, industrial, and residential buildings.

Table 3-6. Accuracy assessment of building change detection.

Area (m^2)	Type	Commercial building	Industrial building	Residential building	Overall
>20m^2	Reference	63	94	177	334
	Correct detection	60	85	168	313
	Wrong detection	14	7	26	47
	Missing detection	3	9	9	21
	Precision	81.08%	92.39%	86.60%	86.94%
	Recall	95.24%	90.43%	94.92%	93.71%
	F1 score	87.59%	91.40%	90.57%	90.20%
>100m^2	Reference	55	70	148	273
	Correct detection	53	65	144	262
	Wrong detection	8	4	11	23
	Missing detection	2	5	4	11
	Precision	86.89%	94.20%	92.90%	91.93%
	Recall	96.36%	92.86%	97.30%	95.97%
	F1 score	91.38%	93.53%	95.05%	93.91%

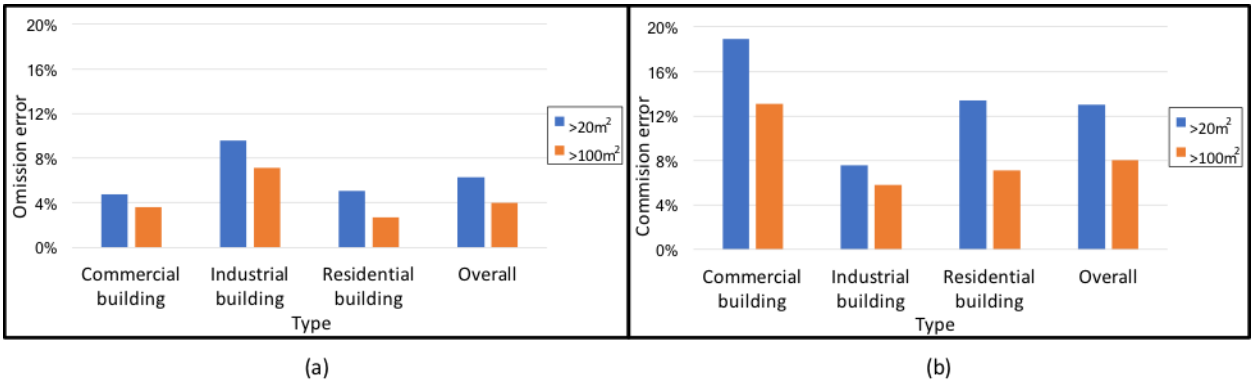


Figure 3-8. Omission and commission errors of different building types.

3.4.4 Visualization of different change types and wrong detections

Visualization is effective to show the four different change types (Figure 3-9). The first and second columns display the 3D points of a changed location in 2002 and 2014 respectively. The third and fourth columns show the corresponding multi-temporal imagery and the extracted building footprints.

The overall omission and commission errors of the change detection results are 6.29% and 13.06% for buildings with the area of footprint larger than 20m^2, and 4.03% and 8.07% for those larger than 100m^2 (Figure 3-8). The first row of Figure 3-10 shows a typical omission error case and the second, third, and fourth rows show three typical commission errors cases. Case one shows an irregular shaped building that the building extraction algorithm fails to extract and recognize the change. Case two shows a building that has some internal change without any obvious volume or footprint change, which causes the method wrongly recognizes it as no change. Case three shows a wrongly detected building change in the downtown area that has shadow issues in 2002 data caused by large scan angles and high-rise buildings nearby. Case four is a structurally complicated building that makes the method wrongly recognize it as a

change because of measurement inaccuracy and data quality variation of the 2002 (poorer) and 2014 (better) LiDAR datasets.

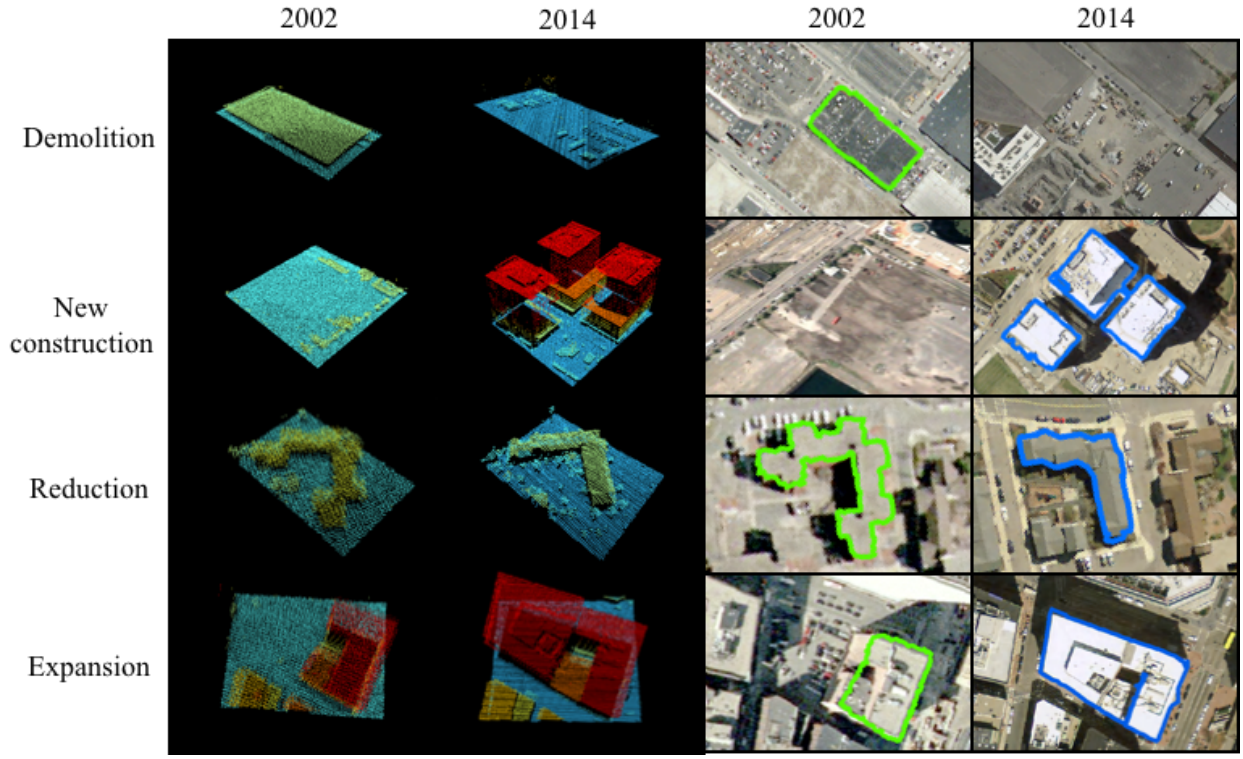


Figure 3-9. Different building change types (green: detected 2002 building footprint; blue: detected 2014 building footprint).

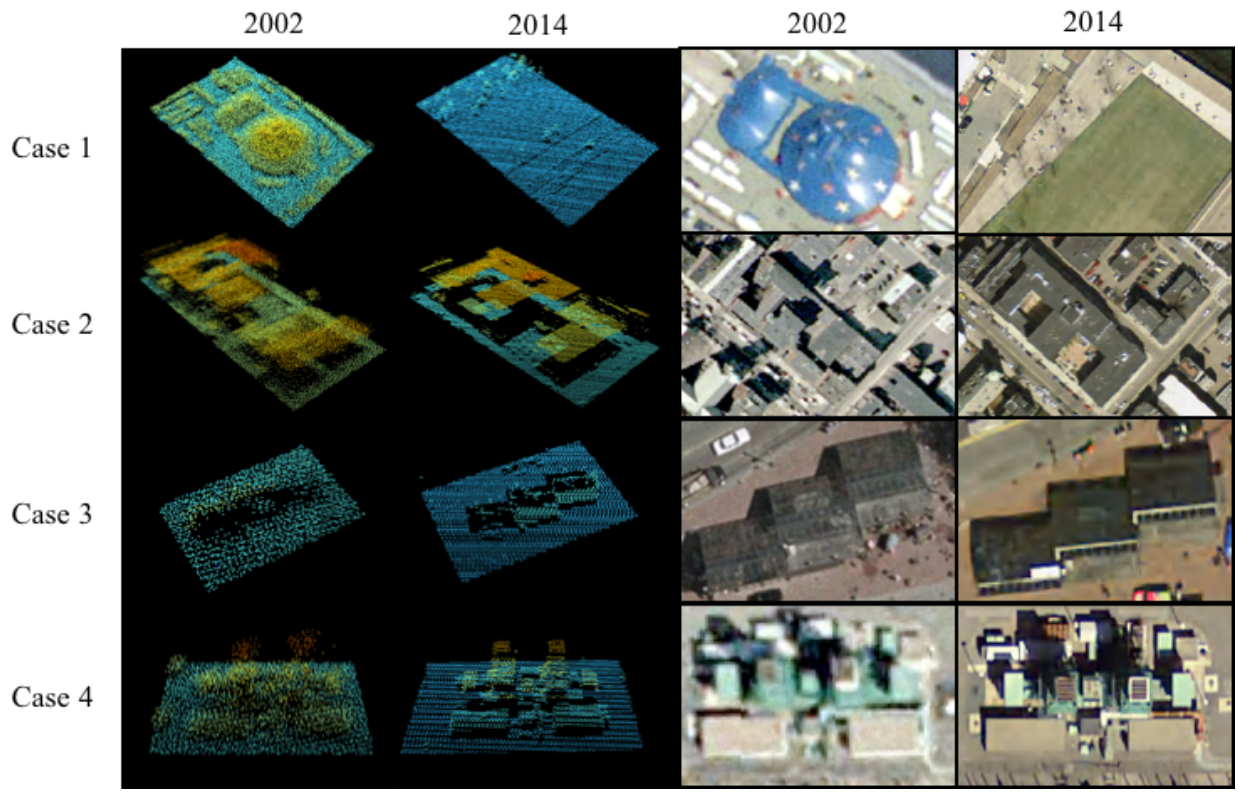


Figure 3-10. Erroneous cases of building change detection.

3.5 Discussions

While urban building extraction and change detection have long been an important research topic, limited research has been done to monitor building changes at individual building level with multitemporal LiDAR data that have become increasingly available. Our research fills this gap by developing a framework using deep learning of a state-of-the-art point labeling algorithm (PointSIFT) and a rule-based method for individual-building based change detection using multitemporal LiDAR data. As the first step, reference data generated from raw LiDAR point cloud is used for the classification of building and non-building points. Then we use Euclidean clustering to isolate individual buildings, register building pairs, calculate the 3D alpha shape of each building pairs, estimate building volume, and finally extract changed

buildings. For accuracy assessment, we use reference data generated by visual interpretation based on a set of multitemporal orthophotography and LiDAR; and overall recall, precision, and F1 score are calculated and organized for the three different building areas. The overall recall, precision, and F1 score of 2002 building classification separately reaches 91.79%, 98.64%, and 95.09% for buildings with footprint areas larger than $20m^2$ and 94.69%, 99.61%, and 97.09% for buildings with footprint areas larger than $100m^2$. The overall recall, precision, and F1 score of 2014 building classification separately reaches 94.55%, 99.38%, and 96.91% for buildings with footprint areas larger than $20m^2$ and 98.87%, 99.66%, and 99.26% for buildings with footprint areas larger than $100m^2$. For change detection, our method successfully extracts most changes by achieving decent overall recall, precision, and F1 score as 93.71%, 86.94%, and 90.20% for buildings with footprint areas larger than $20m^2$ and 95.97%, 91.93%, and 93.91% for buildings with footprint area larger than $100m^2$.

The high classification accuracy proves the effectiveness of the method for the classification of various types of urban buildings at individual building level even with the datasets of limited point accuracy, density, and a low horizontal and vertical accuracy. From the change detection perspective, our method is able to not only avoid detecting changes of non-building objects and wrong or incomplete detection of building changes, but also achieve a high level of completeness and correctness, and reliable and accurate estimation of different types of urban changes. In previous studies, point cloud registration is often the first step for conducting change detection. In our method, we eliminate this computationally intensive process by directly working on individual buildings, where the footprints of individual buildings are delineated and overlapping ones are formed into pairs for comparison of changes. Our method conducts change detection by estimating the change of building volume and footprint, which provides a new way

for quantitatively estimating the amount of building change volume at the level of individual buildings.

Although our method is designed to be generalizable and has weak dependence on high point density and accuracy of LiDAR, there are still several aspects that need to be considered in order to extend it to broader applications. First, in order to utilize the deep learning model, a representative collection of training samples need to be acquired by conducting a point-level labeling. This can sometimes be challenging for locations where points from different classes closely mingle together. Second, an optimal combination of patch sizes for training data needs to be determined based on the building sizes of a study area. Third, the deep-learning algorithm is computationally intensive. Thus, GPU was used in this research to speed up the training process that took about 50 hours.

Based on the workflow (Figure 3-4), the errors of our method can be summarized from three sources. The first one is from the building classification process. The method misses a number of temporarily constructed prefabricated houses that are similar (size, shape, and color) to the large containers (non-building) in the residential and industrial area. The method also fails to recognize some irregularly shaped buildings in the commercial area. Because of the abundant noise in the 2002 data, the algorithm failed to isolate a few adjacent buildings in the residential area. The second is from the change detection process. The method makes some false change detections on several small buildings near high-rise ones, where their points are partially blocked and vary due to different flight angles. Also, some non-building objects closely attached to buildings that change over time are wrongly recognized as changes. The third aspect comes from the data quality. The data collection standards and flight parameters have a large variation

between 2002 and 2014, where the 2002 LiDAR dataset has a much lower accuracy and higher noise rate compared to the 2014 one. This discrepancy leads to part of the false cases.

According to the error assessment, we need to further refine the method to make it more resilient to the aforementioned error sources. Specifically, training data needs to be improved by including additional building types, such as prefabricated houses and irregularly shaped buildings. Regarding change detection, our method is built on the pair-wise comparison of building volumes and footprints. In future, we plan to study pair-wise individual building registration in order to extract more details of the changed buildings. Furthermore, we are committed to testing our method over other study sites and various qualities of datasets for sensitivity analysis to verify its stability and generality.

3.6 Conclusion

Multitemporal LiDAR data have been widely utilized in 3D urban modeling and change detection research. However, limited research has been conducted to examine individual building level changes that are important to urban dynamics research at fine scale. Previous related studies utilize high quality multitemporal LiDAR (+20 pts/m^2) collected for small areas, and built their models on such specific data. This practice has inherent limitations for large geographic areas especially considering that large coverage LiDAR is often available at a much coarser resolution with inconsistent accuracy. This research therefore develops a deep learning method for urban building extraction and change detection at the level of individual buildings using the multitemporal LiDAR data that are readily available with a relatively coarse point density (3 pts/m^2) and a large accuracy difference. Our method also enables the elimination of computationally intensive LiDAR data registration. The evaluation and validation using

multitemporal orthophotography and LiDAR show that the method can achieve high-accuracy individual building extraction and change detection results based on 3D building volume and footprint analysis. The accuracy assessment demonstrates that our method is able to not only achieve high accuracy of individual building extraction, but also provide a high completeness, correctness, reliable and accurate estimation of different types of urban changes.

CHAPTER 4: HYDROLOGICAL STREAMLINE DETECTION USING A U-NET MODEL

Abstract

Surface water is an irreplaceable resource for human survival and environmental sustainability. Accurate, finely-detailed cartographic representations of hydrologic streamlines are critically important in various scientific domains, such as assessing the quantity and quality of present and future water resources, modeling climate changes, evaluating agricultural suitability, mapping flood inundation, and monitoring environmental changes. Conventional approaches to detecting such streamlines cannot adequately incorporate information from the complex three-dimensional (3D) environment of streams and land surface features. Such information is vital to accurately delineate streamlines. In recent years, high accuracy lidar data has become increasingly available for deriving both 3D information and terrestrial surface reflectance. This study develops an attention U-net model to take advantage of high-accuracy lidar data for finely-detailed streamline detection and evaluates model results against a baseline of multiple traditional machine learning methods. The evaluation shows that the attention U-net model outperforms the best baseline machine learning method by an average F1-score of 11.25% and achieves significantly better smoothness and connectivity between classified streamline channels. These findings suggest that our deep learning approach can harness high-accuracy lidar data for fine-scale hydrologic streamline detection, and in turn produce desirable benefits for many scientific domains.

4.1 Introduction

Water is a life-sustaining limited resource on Earth. Less than 0.01 percent of the Earth's water is freshwater stored in lakes and streams that is easily accessible for human use (Shiklomanov, 1991). Growing population and resource consumption have spurred a nearly six-fold increase in global use of freshwater since 1900 (Flörke et al., 2013). Interactions of water within Earth's systems have been studied extensively, yet increased demand for this vital resource has expanded interest in monitoring and management of water resources. Accurate, finely-detailed delineation of surface hydrologic features is crucial for various scientific investigations and water resource applications, such as agricultural suitability, river dynamics, flood mapping, landslide risk analysis, wetland inventory, watershed analysis, environmental monitoring, and climate modeling, to name just a few (Maidment, 2017; Poppenga and Gesch, 2013; Schultz et al., 2017; Simley and Carswell, 2009; Terziotti, 2018; Wright and Nielsen, 2012). The spatial pattern of a surface water drainage network is largely a reflection of the type and arrangement of subsurface bedrock, which can assist with classification and management of land resources (Clubb and Bookhagen, 2019; Muller and Oberlander, 1976). Therefore, the key objective of this research is to understand how to advance machine intelligence for automatic extraction of detailed hydrologic features from high-resolution elevation data and other open geospatial datasets, yielding important data that can be used for this type of scientific work.

The National Hydrography Dataset (NHD) is a digital database of surface water features of the United States that is managed by the U.S. Geological Survey (USGS) and partner organizations (Sheng and Wilson, 2007; Simley and Carswell, 2009). It provides a common reference for regulation, research and modeling (NOAA, 2016). The NHD High Resolution (HR) is a multi-scale dataset comprised from the best available data sources having scales of 1:24,000

or larger (finer detail), except in Alaska where 1:63,360 or finer scales are used. However, the quality of hydrographic data that has been compiled from topographic maps, which include the NHD, is not suitable for certain hydrologic, regulatory, and engineering purposes because of inconsistent drainage density and some missing headwater content (Caruso, 2014; Chorley and Dale, 1972; Colson and Gregory, 2008; Colson, 2006; Fritz et al., 2013; Russell, 2008). Headwaters are small streams formed at the upstream extent of a watershed and comprise more than 50 percent of the stream network by length in the United States (Nadeau and Rains, 2007). To overcome these issues, since 2009 NHD HR is being updated with more detailed hydrography derived from finer-scale source information, up to 1:2,400 (Simley and Carswell, 2009; Stanislawski, 2009). Economically, these enhanced hydrographic data are expected to generate over 600 million dollars per year in potential benefits, in addition to the 500 million dollars in annual benefits already being generated from the existing program (Hoegberg, 2016).

Updating the NHD HR applies the best available digital elevation model (DEM) data, which should use Quality Level 2 (QL2) LiDAR data or better. Since 2014, the USGS 3D Elevation Program (3DEP) has been coordinating the collection of QL2 or better LiDAR point cloud data for the United States, except for Alaska where cloud-penetrating interferometric synthetic aperture radar (ifsar) is being acquired to simplify collection in remote areas (Lukas et al., 2015). QL2 LiDAR provides a nominal pulse spacing of less than 1-meter (m) for ground return points (Heidemann, 2018), which adequately supports derivation of a 1-m resolution DEM. The detail inherent to this high-resolution DEM data enables modeling of surface water dynamics from the continental scale to the headwater and catchment scales.

Although some methods to improve the NHD HR have been studied (Lopez-Torrijos, 2018; Poppenga et al., 2013; Sheng et al., 2007; Stanislawski and Survila, 2018), extracting

accurate and fine-scale hydrography from high-resolution DEM data using traditional flow accumulation methods is a costly and laborious process. Depending on the selected workflow, various sophisticated issues must be handled, which include conditioning the DEM for flow modeling, estimating flow accumulation weights and a minimum contributing area for stream formation, along with tailoring solutions to diverse environmental conditions. Coupled with the fact that multiple methods are available, solutions can vary, and assessing the accuracy of extracted drainage lines is further complicated by temporal environmental variations. Procedures generally involve well-known automated methods to derive drainage lines from DEMs (Anderson, 2012; Jenson and Domingue, 1988; Maidment and Morehouse, 2002; Metz and Mitasova, 2011; Montgomery and Foufoula-Georgiou, 1993; O'Callaghan and Mark, 1984; Passalacqua and Belmont, 2012; Poppenga et al., 2013; Tarboton and Bras, 1991), with subsequent manual editing to adjust drainage lines and collect waterbodies from high-resolution orthorectified images.

Remotely sensed information at high spatial and temporal resolutions, such as LiDAR big data, can facilitate automated analysis and extraction of hydrographic features, saving time and increasing the accuracy and consistency of extracted features (Sharma and Xu, 2016). Advanced computationally-intensive machine learning approaches integrated with cyberGIS (cyber geospatial information science and systems) for the resolution of computational and data intensive geospatial analysis (Wang, 2010; Wang and Goodchild, 2019; Wang and Liu, 2016), represent an exciting frontier for extracting accurate and fine-scale hydrography from LiDAR to improve the NHD HR.

Recent rapid advances in deep learning have been widely acknowledged and adopted in many challenging pattern recognition and object detection tasks (Kampffmeyer and Salberg,

2016; LeCun and Bengio, 2015; Maggiori and Tarabalka, 2017; Reichstein et al., 2019; Schmidhuber, 2015; Sun and Zhang, 2018; Xu and Guan, 2018; Zhu et al., 2017). Compared to the traditional or hand-crafted feature engineering, deep learning has demonstrated advances in accuracy and efficiency for complex feature learning in various application domains (Liang and Sun, 2017; Lin and Tegmark, 2017; Lin and Nie, 2017; Xu and Mountrakis, 2017). While such strategies promise a new way for hydrologic feature extraction from geospatial big data, limited effort has taken advantage of deep learning for accurate, efficient, and fine-scale delineation of hydrologic features. Moreover, the full utilization of the most recent technology of Geiger-mode LiDAR could significantly improve high-quality delineation of natural features (Clifton et al., 2015; Stoker and Abdullah, 2016).

This research develops a deep learning model based on the U-net structure (Ronneberger and Fischer, 2015) and attention mechanism (Oktay et al., 2018; Vaswani et al., 2017), which consists of a contractive path and an expanding path for segmenting streamlines from input feature maps. The contractive path is comprised of six triple convolutional layers plus five pooling layers for accurate extraction of global features and reduction of spatial redundancy, while the expanding path is comprised of five transposed convolutional layers plus five triple convolutional layers for projecting the extracted global feature content to original locations in the prediction map. The number of layers is chosen to reduce (contractive path) and upsample (expanding path) the x-y dimension of feature maps from 224-by-224 to 7-by-7 or from 7-by-7 to 224-by-224 with a stride of two. The patch size of 224 influences the model accuracy and efficiency. Small patch sizes cause poor accuracies due to the lack of context information while large patch sizes add extra computational burden without particular benefits to model performance. In this research, patch sizes of 64, 112, 224, 448, 512 were tested and 224 was

chosen over smaller or larger ones based on evaluation of model accuracy and efficiency. Meanwhile, feature concatenation is used to combine the extracted local and global information at different levels from the contractive path to its corresponding locations in the expanding path, to enhance the expressivity of the model during the convolution and transposed convolutional processes.

Computationally, we use GPU processing to speed up model training which is based on Keras and TensorFlow. Two types of benchmark methods are adopted for model comparisons. The first type includes two traditional pixel-based classification methods – a Support Vector Machine (SVM) and an Artificial Neural Network (ANN) model. The other types include the NHD HR data compiled from topographic maps and orthophotography, and elevation-derived drainage lines generated from Geonet tools (Sangireddy and Stark, 2016). The comparison shows that our method based on the attention U-net model outperforms the best benchmark method by 8.61%, 9.39%, 13.68%, and 13.31% in four different scenarios. The resulting streamline map also indicates that the attention U-net model generates smoother and more topologically connected features than the benchmark methods, which is significant for hydrologic applications. The major contributions of this research are two-fold: (1) a novel application of the attention U-net for accurate and fine-scale hydrologic streamline detection, and (2) an effective streamline detection method that fully utilizes the geometric and intensity information from high-resolution LiDAR data.

4.2 Study area and dataset

Our study area is a watershed in Rowan county, which is located in central western North Carolina (Figure 4-1). This area encompasses a set of tributaries that flow into Second Creek,

which is the primary flowline feature of 12-digit NHD watershed 030401020504. The study area is 6.07 km2 and has a humid, subtropical climate. Winters are short and mild, while summers are usually sultry. Spring and fall are distinct and refreshing periods of transition. Temperature ranges between 100F (38C) to 10F (-12C) within a year. It lies in the Central Interior and Appalachian ecological division (Comer et al., 2003). In terms of land cover types, forest dominates a majority of the area and most of the stream channels are underneath closed-canopy forest. The LiDAR dataset used in this research is small-footprint, discrete-return, Geiger-mode LiDAR that was collected by the state government of North Carolina in the fall of 2016. The LiDAR dataset requires about 21 GB of disk storage, and the projection coordinate system is the 2011 State Plane of North Carolina. This area has elevations ranging from 669 to 820 meters. Because a Geiger-mode LiDAR sensor was used, the point density of all returns reaches 43 returns per square meter. A field validated set of intermittent stream heads surveyed between 2013 and 2014 along with on-screen editing is used to generate reference data (Shavers and Stanislawski, 2018). A 3-meter buffer is generated along the reference streamline in order to take into account the width of stream channels.

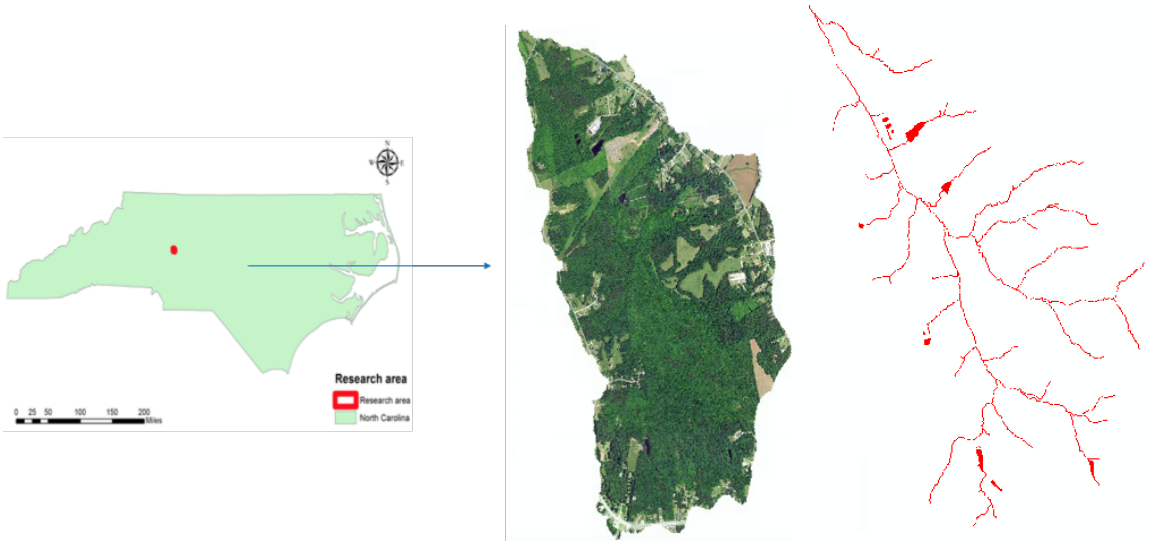


Figure 4-1. Research area (Left: boundary of North Carolina state; middle: a 1-meter resolution image from the National Agriculture Imagery Program; right: reference data).

Eight co-registered 1-m resolution raster data layers were derived from the LiDAR point cloud data and were used for training, validation, and testing in the research. The layers were selected through extensive comparison of elevation derivatives and optical imagery having national coverage with validated surface hydrography in diverse landscapes. The raster layers include: (1) a 1-m resolution digital elevation model (DEM) derived from the ground return points; (2) geometric curvature determined from the DEM; (3) a topographic position index (TPI) derived from the DEM using a 3-cell by 3-cell window; (4) a TPI derived from the DEM using a 21-cell by 21-cell window; (5) zenith angle openness derived from the DEM using a 10-cell radius with 32 directions (Doneus, 2013); (6) return intensity determined from the LiDAR ground points averaged with inverse distance weighting using 10 nearest points; (7) point density for return points between zero and 1 foot above ground; and (8) point density for return points between 0 and 3 feet above ground. Geometric curvature is determined using GeoNet software (Sangireddy et al., 2016). The software applies the non-linear diffusion Perona-Malik filter on the DEM to remove noise and sharpen the localization of channels (Passalacqua and Do Trung,

2010). Geometric curvature, which sums curvature in the x and y directions, is then determined for the filtered DEM. The TPI value of a cell is the difference between the cell elevation and the local average elevation within a specific radius or within a surrounding window of cells (De Reu et al., 2013). As noted, average values for TPI layer (3) and (4) are computed based on 3x3 and 21x21 surrounding cell windows, respectively. The TPI exaggerates local lows and highs in a DEM relative to the nearby topographic features, accentuating ridges and valleys. Return point density layers (7) and (8) respectively estimate the density of land surface features up such as shrubs and tree limbs to 1 and 3 feet above ground, which is most likely vegetation under the forest canopy. Shavers and Stanislawski (2018) suggest vegetation density structure in the riparian zones may be reflected in these layers. The eight raster layers are shown in Figure 4-2 with summary statistics presented in Table 4-1.

Inputs to our model are individual image patches sampled from the eight different feature maps derived from the LiDAR data at a resolution of 1m. The feature maps are normalized versions of the eight raster data layers, normalizing each of the floating point datasets within the study area to a corresponding unsigned integer feature map. In order to effectively test our method, we created four different classification scenarios by splitting the research area into upper/lower and left/right portions. When conducting our experiments, we used one of the portions for generating training/validation patches, and the other portion to generate testing patches for accuracy assessment in order to evaluate model generalizability. Sample patches for training and validation are generated based on a random process that ensures no overlap between training and validation patches. A visualization of the locations of our generated training and validation patches is shown in Figure 4-3. To further enhance our training data, we apply image augmentation by randomly rotating each training patch by 30-150 and 210-330 degrees,

rescaling each sample by 0.5-0.8 and 1.5-2.0, shearing each sample by random ranges from -30 to 30 degrees, and mirroring each sample horizontally to finally create six augmented samples for each training sample. Finally, 200 training (1400 after data augmentation) and 30 validation patches are selected.

Table 4-1. Summary statistics for 1-meter resolution raster datasets derived from LiDAR point cloud data collected for Rowan County watershed.

Raster Dataset	Minimum	Maximum	Mean	Standard Deviation	Range
Digital elevation model (meters)	669.95	820.13	752.75	37.96	150.18
Geometric curvature	-97.25	97.93	0.01	3.05	195.18
Topographic position index (3x3 window)	-8.59	5.58	6.38	0.18	14.167
Topographic position index (21x21 window)	-13.62	13.29	0	0.93	26.91
Openness (R10, D32) degrees	21.52	118.8	83.41	7.35	97.28
Return intensity	0	55185.39	29047.18	10624.11	55185.39
Return point density 1 ft above ground (points per m^2)	0	0.94	0.02	0.04	0.94
Return point density 3 ft above ground (points per m^2)	0	2.89	0.12	0.23	2.89

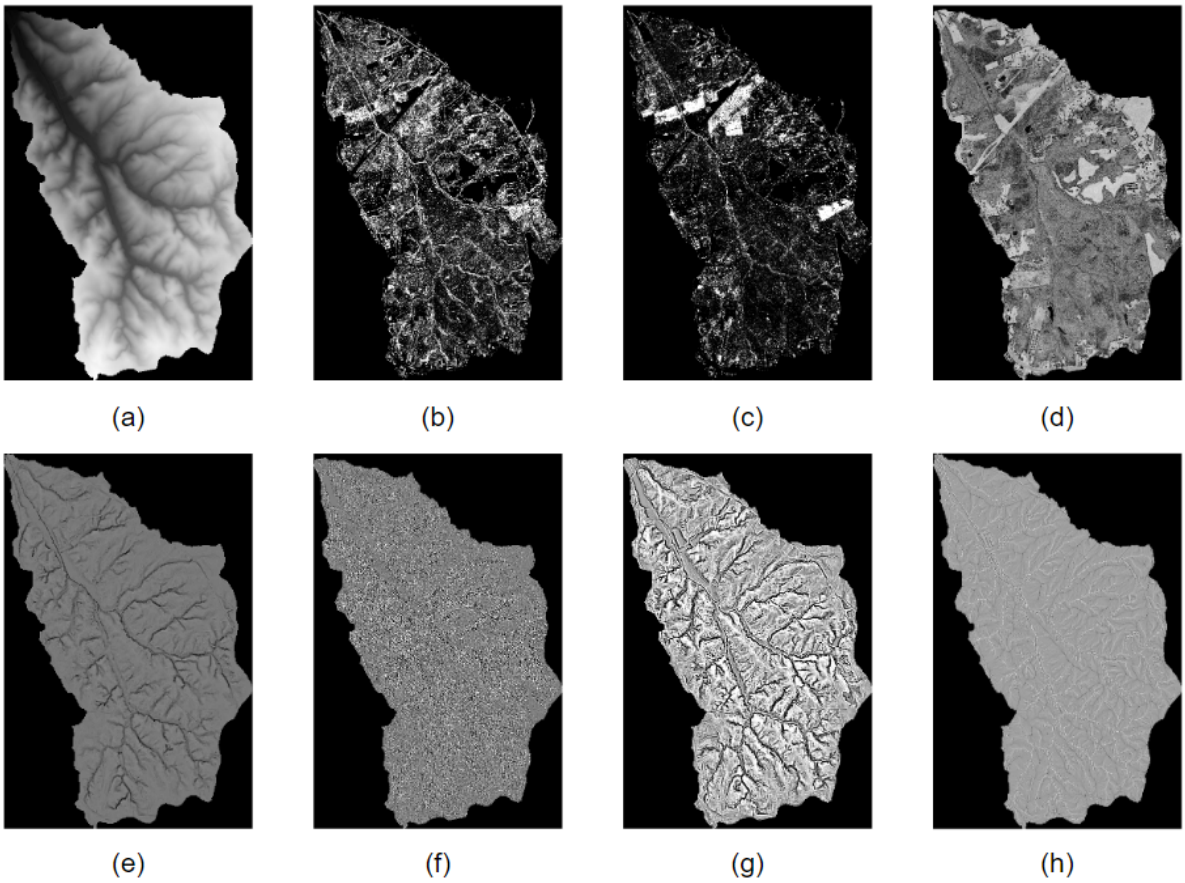


Figure 4-2. LiDAR feature maps: (a) Digital Elevation Model (DEM); (b) Digital Surface Model (DSM) 1 feet below; (c) Digital Surface Model (DSM) 3 feet below; (d) LiDAR Reflectance; (e) Geomorphon openness; (f) TPI with moving window size 3; (g) TPI with moving window size 21; (h) Topological Curvature.

Figure 4-3. Distribution of training and validation patches in the four scenarios (gray area: locations of training patches; white area: locations of validation patches).

4.3 Methods

4.3.1 Benchmark methods

To evaluate the performance of our method, it is compared to existing hydrography data, elevation-derived drainage lines, and hydrography predicted from two machine learning methods. We set up four baseline benchmarks, which include NHD HR, elevation-derived drainage lines from Geonet, Support Vector Machine (SVM), and an Artificial Neural Network (ANN) model. The high resolution NHD HR was retrieved from the USGS NHD website (https://www.usgs.gov/core-science-systems/ngp/national-hydrography). NHD data for the study area were compiled from 1:24,000-scale digital line graph data in 2001, with waterbodies and associated features manually adjusted in 2013 to fit National Agriculture Imagery Program (NAIP) 1-m resolution color-infrared digital orthophotography. The Geonet lines are extracted from the DEM using a least-cost path-tracing technique that is guided by a minimum threshold flow accumulation skeleton (Sangireddy et al., 2016). In our case, the flow accumulation skeleton is generated using a minimum threshold of 1000 cells, which is expected to over extract water flow network and fully define the drainage paths.

We use an SVM classifier based on a radial basis function (RBF) kernel with a kernel approximation strategy for speeding up the training process (Rahimi and Recht, 2008). The parameters of kernel degree (g) and penalty (C) are tuned using a two-level grid search in the range of 10-5 to 105 and 10-5 to 1, respectively. For the benchmark neural network model, we construct a model with two hidden layers and a sigmoid activation function as the output layer. The parameters of number of hidden layers, learning rate, momentum, decaying rate are also tuned using grid search. The reference data including training, validation, and testing data are the same between different models for model training, parameter tuning and generating the final feature maps.

4.3.2 The U-net model

The U-net model is a special type of Fully Convolutional Networks (FCNs). Unlike normal CNNs, the last fully connected layer from FCNs are substituted by a series of transposed convolutional layers with larger and larger receptive fields. FCNs are built only by locally connected layers including convolution, pooling and upsampling layers without using any dense connected layer. This practice greatly reduces the number of parameters for model tuning and thus reduces redundant computation compared to traditional convolutional neural networks. A typical FCN has two parts: a contractive path and an upsampling path, where the former is used to extract important information and reduce spatial redundancy, and the latter is used to project the extracted information to specific locations in the original image (Ronneberger et al., 2015). The U-net model is a state-of-the-art FCN that achieves high accuracy for solving image segmentation problems (Ronneberger et al., 2015). Based on the fundamental structure of an FCN, it further applies feature concatenations to recover and fully utilize the information

extracted at different resolution levels in the contractive path to the corresponding locations in the expanding path.

Convolutional layer: the convolutional layer is the major workforce for extracting important features from images (Krizhevsky and Sutskever, 2012). It conducts image filtering by using kernel filters. In this process, image features with strong signals are extracted.

Pooling layer: Pooling layer is used to conduct downsampling on activation maps. Downsampling reduces the sampling rate of a raster by decreasing the raster resolution (i.e., increasing pixel size). The max function is often used to filter out redundant information and preserve the strongest feature signals.

Relu layer: Relu layer is short for rectified linear unit layer, which is one of the most commonly used activation functions in convolutional neural networks (Agarap, 2018). It consists of a linear function for all positive input values, and zero for all negative values. It truncates unimportant features generated from the convolutional layer and only reserves the important ones.

Transposed convolutional layer: This layer projects the extracted dense features from the coarse resolution to its precise location in the original image by using upsampling (i.e., increasing pixel resolution) or spatial interpolation.

The attention module is a technique that is originally designed for sequence dependency modeling, and has recently been adopted for modeling feature dependencies in image analysis (Oktay et al., 2018; Vaswani et al., 2017). It can progressively suppress feature responses in irrelevant background regions and make the model focus on important features. In this research, we integrate five attention gates (AGs) into the U-net model and thus create an attention U-net model for achieving high accuracy results. As shown in Figure 4-4a, the vanilla attention module

maps query pixels and their key-value pairs to the output. The output is a set of weighted values and the attention weight matrix is calculated by a compatibility function of the query with the corresponding key (Vaswani et al., 2017). Finally, the weighted inputs are multiplied by a scaling hyperparameter α (initialized as 1) and added to the original input to produce the final output. Since the original attention weight matrixes at shallower layers are too large to fit in the memory, we used a second attention module (Figure 4-4b) adapted from (Oktay et al., 2018) in our model. The only difference is the second one directly combines the convolutional results from the feature maps and the gating signal to a Relu layer to remove negative values, and utilizes a bottleneck convolutional layer to reduce the channel dimension for memory saving and a sigmoid function to calculate the final attention weight matrix.

The architecture of the attention U-net model is shown in Figure 4-5, it applies six triple convolutional layers to the contractive path and five in the expanding path. Five pooling layers are used between each of the triple convolutional layers for downsampling. In the expanding path, five transposed convolutional layers (size 2×2 and stride 2) are used for feature upsampling. In each horizontal level of the two paths, the network uses attention gates to filter the features propagated through the skip connections based on the gating signal of the contextual information from coarser scales to achieve high accuracy in segmentation results. We utilize an "Adam" optimizer for calculating the change direction of loss and adjust the weights in the back-propagation process (Kingma and Ba, 2014).

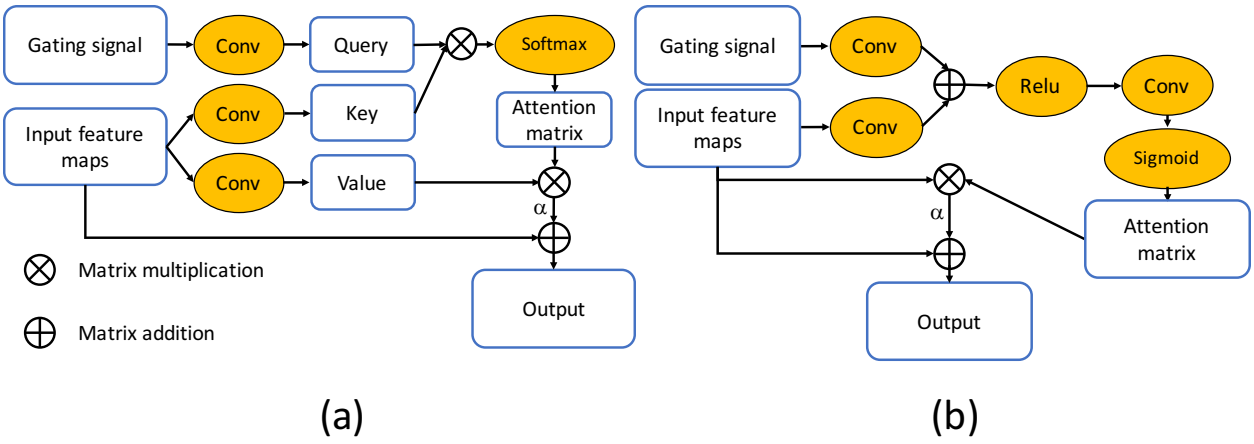


Figure 4-4. Attention gates of the U-net model (a: attention gate 1; b: attention gate 2).

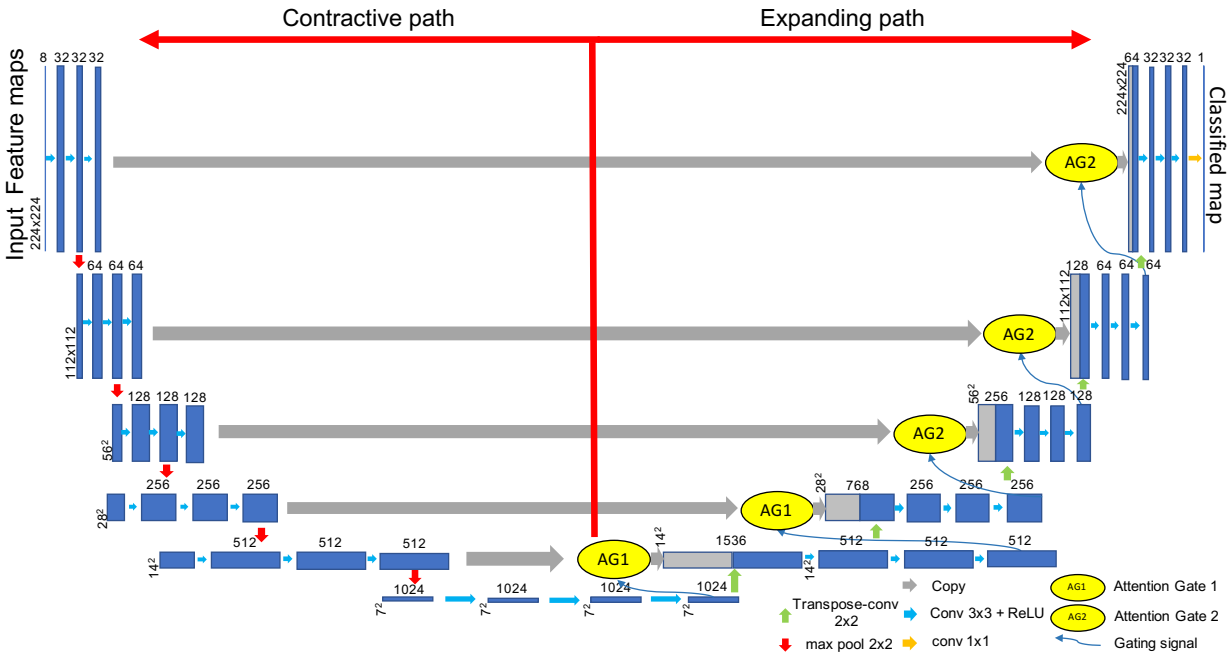


Figure 4-5. The architecture of the attention U-net model.

We use Python 2.7 and keras 2.0 with backend of tensorflow 1.0 for the model construction. We also utilize python libraries including sklearn 0.18.1, scikit-image 0.16.2, gdal 3.0.2, numpy 1.17.3 supported by Anaconda 2.0. The model is tested using both GPU and CPU devices. It takes 15 hours using a state-of-the-art CPU and 2 hours for a Tesla M80 GPU to

finish model training for our 6.07 km2 study area. In this paper, we run the attention U-net and the U-net model separately five times and the average statistics are reported for evaluation.

The attention U-net model uses Dice's coefficient (4.1) as the loss function. The coefficient is the quotient of similarity and ranges between 0 and 1. Dice's coefficient value equals twice true positive (TP) divided by the sum of twice true positive (TP), false positive (FP), and false negative (FN) as shown in the equation (4.1). We use precision, recall, and F1 score to evaluate the model performance against testing data. Because of the difficulty in correctly labeling all streamline pixels, relaxed methods are adopted to calculate precisions and recalls (Mnih and Hinton, 2010). The relaxed precision is defined as the fraction of number of pixels predicted as stream within a range of ρ pixels from pixels labeled as stream. The relaxed recall is the fraction of number of pixels labeled as stream that are within a range of ρ pixels from pixels predicted as stream pixels. In our experiments, the slack parameter ρ is set to 3 according to previous research (Mnih and Hinton, 2010).

$$Dice's\ coefficient = \frac{2\mathrm{TP}}{2TP+FP+FN} \tag{4.1}$$

We used a grid search for hyperparameter tuning of the learning rate, filter size, dropout rate, and decaying factor. In this process, the "Adam" optimizer is used to calculate and adjust the weights during training (Zhang, 2018). Figure 4-6 shows the change of training accuracy from different learning rates against the number of training epochs using vanilla U-net model. We can see that the learning rate of 3.59e-05 achieves a good convergence and accuracy. The plot of training and validation losses using the selected learning rate is shown in Figure 4-7.

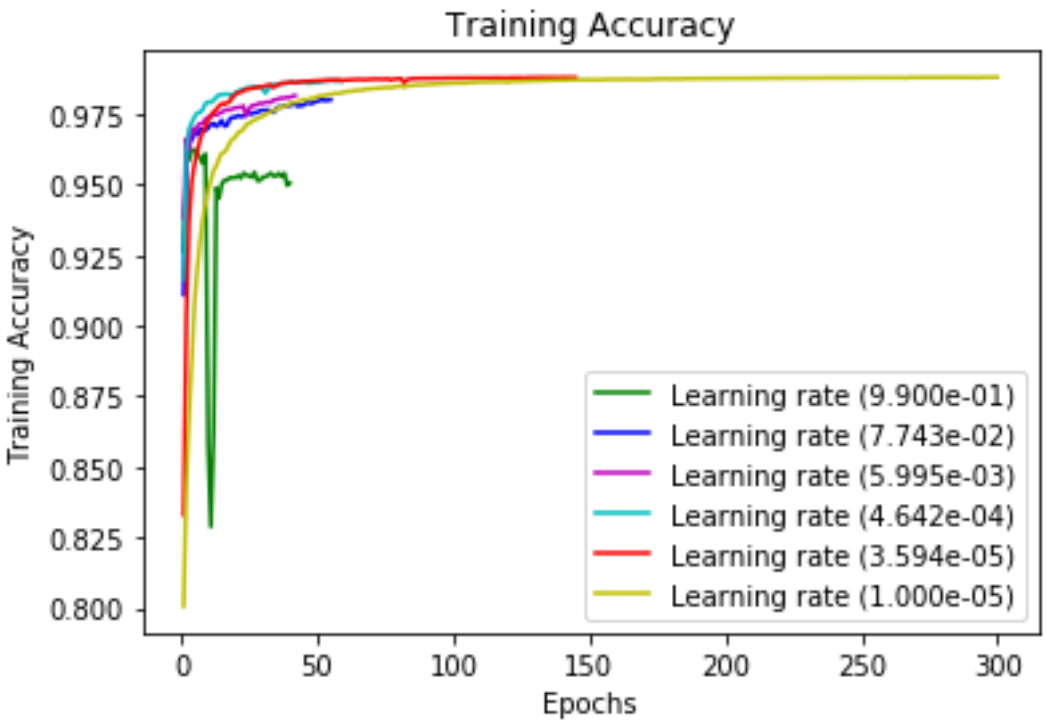


Figure 4-6. Training accuracies with different learning rates.

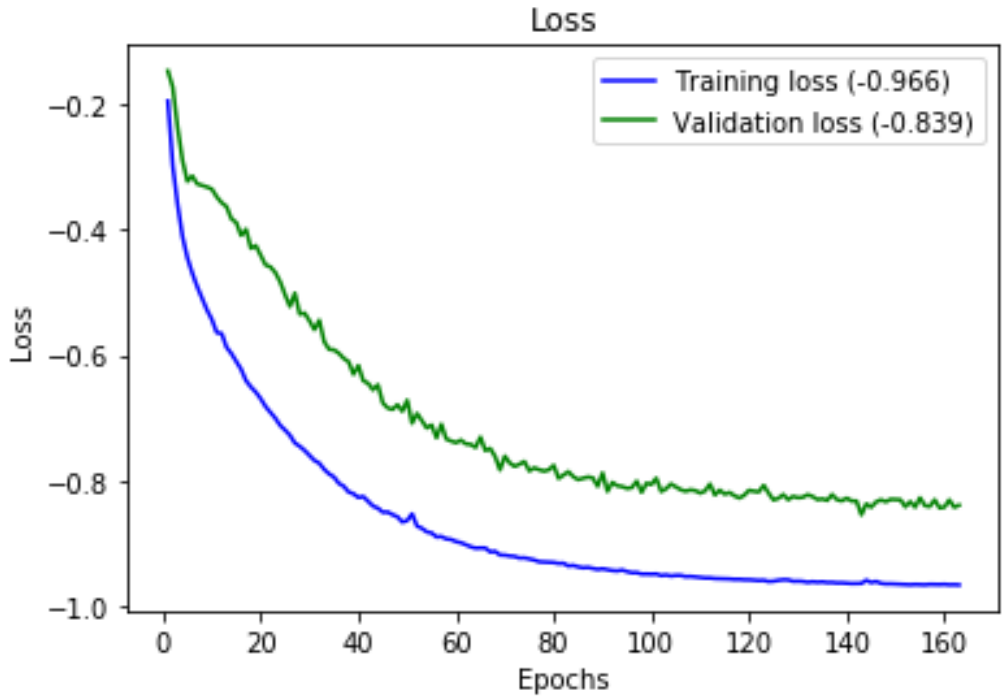


Figure 4-7. Loss of the training and validation against epochs.

4.4 Result

We evaluate our method against multiple benchmark methods in four scenarios. For the first two scenarios, we split the study area horizontally and use the lower portion as testing data in scenario one, and the upper portion as testing data in scenario two. For scenarios three and four, we split the study area vertically and use the right portion as testing data in scenario three, and the left portion as testing data in scenario four. Three metrics, precision (4.2), recall (4.3), and F1 score (4.4) are used to evaluate the performance of the methods and are defined as follows (TP: True Positive; FP: False Positive; FN: False Negative).

$$Precision = \frac{TP}{TP+FP} \tag{4.2}$$

$$Recall = \frac{TP}{TP+FN} \tag{4.3}$$

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4.4}$$

In this research, we focus on the F1 score because it is the harmonic mean of precision and recall. The highest F1 score means the model has an optimal balance of recall and precision. While recall expresses the model's ability to find all streamline pixels in the input data, precision expresses the portion of pixels that a model classifies as streamline correctly. Therefore, there is a trade-off between these two metrics.

The F1 score, precision, and recall of test accuracies for the attention U-net, U-net, and benchmark methods among the four scenarios are shown in Table 4-2, 4-3, and 4-4 respectively. Table 4-2 shows that both the attention U-net and U-net models outperform all of the benchmark methods for the four scenarios in overall F1 score. Also, the attention U-net slightly outperforms the U-net model in terms of the average F1 score. Among the four benchmark methods, ANN achieves the highest accuracy and NHD has the lowest accuracy. The attention U-net model outperforms ANN by 8.61%, 9.39%, 13.68%, and 13.31%; SVM by 12.12%, 12.98%, 22.69%,

and 13.54%; NHD by 45.51% and Geonet by 23.61% on average. For precision, SVM achieves the best and outperforms the attention U-net model by 12.04%, 7.64%, 1.99%, and 9.51% from scenario 1 to 4. The attention U-net model outperforms the U-net model by 4.73%, 3.23%, 3.18%, and 2.84% from scenario 1 to 4. For recall, the Geonet model has the highest accuracy of 92.66%, which is 1.78% higher than the attention U-net model. Apart from that, the attention U-net model achieves the next highest recalls and outperforms the U-net model by 0.64%, SVM by 32.15%, ANN by 17.62%, and NHD by 54.92% on average in scenario 1 to 4. Overall, the attention U-net model outperforms the U-net and all of the benchmark methods according to the average F1-score. Although one of the benchmark methods generates better precision values, errors reflected by recall values are large and makes it worse in terms of general performance compared to both the attention U-net and U-net model.

Table 4-2. Comparison of the F1 scores between the attention U-net, the U-net and benchmarks.

Scenarios (F1 score)	Attention U-net model	U-net model	SVM	ANN	NHD	Geonet
Scenario 1	83.02%	78.94%	70.90%	74.41%	41.05%	62.95%
Scenario 2	90.53%	87.94%	77.55%	81.14%		
Scenario 3	91.91%	90.61%	69.22%	78.23%		
Scenario 4	80.79%	79.82%	67.25%	67.48%		
Average	86.56%	84.33%	71.23%	75.32%		

Table 4-3. Comparison of the precisions between the attention U-net, the U-net and benchmarks.

Scenarios (Precision)	Attention U-net model	U-net model	SVM	ANN	NHD	Geonet
Scenario 1	74.53%	69.80%	86.57%	80.90%	47.84%	47.67%
Scenario 2	87.96%	84.73%	95.60%	82.40%		
Scenario 3	91.06%	87.88%	93.05%	82.34%		
Scenario 4	78.16%	75.32%	87.67%	65.38%		
Average	82.92%	79.43%	90.72%	77.76%		

Table 4-4. Comparison of the recalls between the attention U-net, the U-net and benchmarks.

Scenarios (Recall)	Attention U-net model	U-net model	SVM	ANN	NHD	Geonet
Scenario 1	93.80%	91.10%	60.04%	68.88%	35.96%	92.66%
Scenario 2	93.27%	91.41%	65.24%	79.92%		
Scenario 3	92.78%	93.53%	55.10%	74.50%		
Scenario 4	83.66%	84.93%	54.55%	69.73%		
Average	90.88%	90.24%	58.73%	73.26%		

We visualize two large-extent locations (Figure 4-8) and two further zoomed-in contexts (Figure 4-9) in scenario 1 to demonstrate the improved performance compared with the benchmark methods. Both the attention U-net and U-net model generate better streamline delineations with better connectivity and smoother shapes following channels compared to ANN and SVM, which generates fragmented channels. The NHD vector features are smooth and well-connected, as are the Geonet drainage lines, which are generated from a least-cost path model guided by flow accumulation (Sangireddy et al., 2016). However, Geonet lines overestimate channels by false recognition of the dry drainage lines as stream channels. As expected from 1:24,000-scale data the NHD is sparse and only contains the several major channels in the study area. The U-net and attention U-net model also perform better in extracting most water-related features including water bodies in the two locations, where all the automated benchmark methods fail to do so. NHD includes the small lakes interpreted from orthophotography. When the attention U-net model is compared to the U-net model, the former eliminates many overestimated streamlines in the middle part of location 1 and 2, and better extracts water bodies. From Figure 4-9, we can see that the attention U-net model is superior to traditional machine learning methods in extracting smooth streamlines and water bodies and avoids a majority of overestimated streamline pixels in NHD HR and the Geonet flow accumulation model. The U-

net model performs similarly to the attention U-net model, but the latter has a better delineation of the streamlines in the north and middle parts (less overestimations) of location 1 and better delineation of water bodies of location 2.

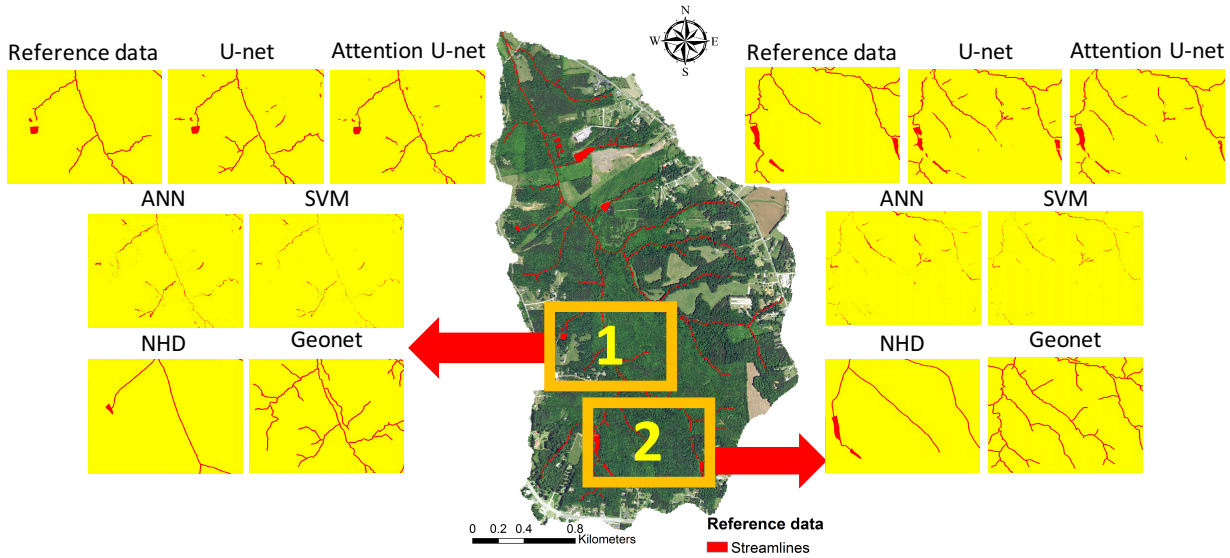


Figure 4-8. Comparison of classification results from the attention U-net, U-net, and different benchmark methods

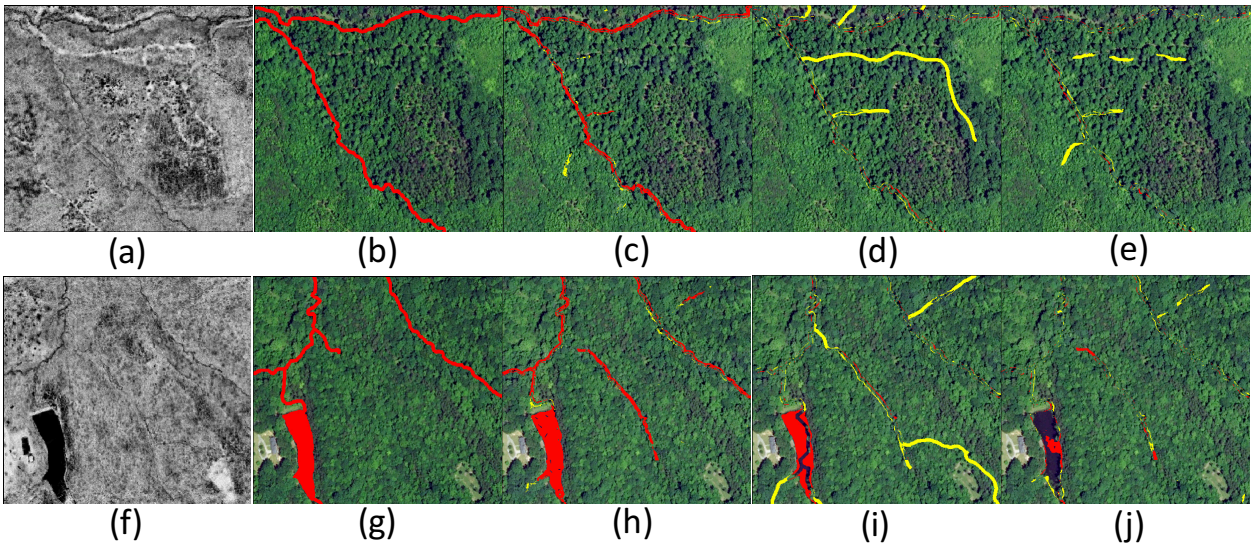


Figure 4-9. Visualization of the feature differences between the attention U-net model versus the Geonet and ANN models for two zoomed-in contexts: (a) and (f): Ground reflectance maps from LiDAR; (b) and (g): the reference data; (c) and (h): the difference map of the attention U-net model and the ANN model; (d) and (i): the difference map of the attention U-net model and the Geonet model; (e) and (j): the difference map of the attention U-net model and the U-net model. Red areas are the pixels labeled as the stream feature by the attention U-net model and recognized as the non-stream feature by the benchmark models or the U-net model. The yellow areas are the pixels labeled as stream feature by the benchmark models or the U-net model and recognized as non-stream feature by the attention U-net model.

4.5 Conclusion and Discussion

This research developed an attention U-net model for hydrologic streamline extraction using LiDAR-derived feature maps. Specifically, we have solved an image segmentation problem (segmentation of streamlines) based on the binary classification of stream versus non-stream pixels. This problem is difficult because hydrologic streamlines are formed by complex processes and occupy only a small portion of diverse land cover types while extracted

streamlines need to be well-connected. Furthermore, surface water features (e.g., clear/turbid rivers, swamps, ponds, and lakes) are spatially heterogeneous, and thus are difficult to extract using traditional machine learning methods (e.g., ANN and SVM) that cannot effectively handle multi-scale context information (e.g., topology, land cover distribution, topography).

The U-net model is a special type of fully convolutional neural networks using skip connections to combine local content from the contractive path to global content in the expanding path, which ensures adequate connectivity of segmentation results because it enables the model to take both of the global and local context information into consideration while extracting streamlines. The attention module is added to the U-net model to progressively suppress feature responses in irrelevant background regions and make the model focus on important streamline features according to the reference data. Addition of the attention module further enhances the accuracy for difficult instances such as the boundary of lakes, river bends, and dried channels.

A comprehensive evaluation of the model shows that our method outperforms multiple machine learning models and conventional flow accumulation methods by providing smoother and better-connected streamline and waterbody features. In order to evaluate the model thoroughly, we created four different scenarios by splitting our research area into upper/lower and left/right portions for generating training/validation and testing patches respectively. The attention U-net model generates F1-scores of 83.02%, 90.53%, 91.91%, and 80.79% across the four scenarios, which outperforms the best benchmark by 8.61%, 9.39%, 13.68%, and 13.31%.

Streamlines extracted using ANN and SVM are fragmented with missing parts mainly because these pixel-based classification methods fail to consider the global context. SVM achieves a high-precision result but a poor recall. This indicates that it underestimates the stream

class pixels but provides good confidence of those extracted. Since the reference data are highly imbalanced (1:100 between stream and non-stream), we also conducted additional experiments that artificially upsample the stream samples. Although the imbalanced issue is resolved, the model has heavy overestimation of the stream class and the accuracy is not comparable to the imbalanced case. Compared to the manually verified reference data, the elevation-derived Geonet drainage lines and features furnish higher recall scores than precision scores, which indicates that this method overestimates the stream class in general and performs better in terms of completeness than precision. These overestimations appear mainly related to local climate variations that cause more dry tributaries and drainage lines than collected in the reference data. The 1:24,000-scale NHD only contains major stream channels and lake features and ignores the smaller tributaries, so the accuracy is much lower than the other datasets. Only the NHD benchmark data includes the water bodies in the study area, and the attention U-net model performs much better in terms of water body extraction than the other benchmark methods.

The attention U-net model utilizes its special feature concatenation design and a fully convolutional neural network to achieve high accuracy, adequate connectivity, and efficient streamline detection. Conventional machine learning models produce less optimal results primarily because they employ a pixel-based classification strategy. The attention U-net model also departs from traditional flow accumulation models that heavily rely on expert inputs, which in this case includes over-extracted drainage lines and no water body extraction leading to large errors. This research utilizes Geiger-mode LiDAR, which provides high-density point clouds and precise measurements for enabling transformative discovery and innovative opportunities in many scientific domains (McManamon et al., 2017).

The following set of principles distilled from this research are important for guiding the application of the method to other areas of study or solving similar problems.

- Ensure a balanced number of convolutional layers at each horizontal level in both the contractive and expanding paths. We find triple convolutional layers achieved adequate results and adding more convolutional layers would not benefit the model but increase computational intensity.
- Training patches should be randomly generated and have no overlap with validation and testing data. Training patches can overlap with themselves (effect of data augmentation).
- Use data augmentation by randomly rotating, mirroring, shearing, and rescaling training samples to ensure the expressivity of the model.
- A dropout layer and a proper dropping rate (hyperparameter) are necessary in the final convolutional process for the model regularization.
- Use early stopping to enhance the training efficiency and prevent overfitting of the model.

Future work will focus on applying the method to more study areas, and scale up the model to regional and national scopes.

CHAPTER 5: AN INTEGRATED CYBERGIS AND DEEP-LEARNING FRAMEWORK FOR SCALABLE LAND COVER MAPPING USING LIDAR AND LANDSAT REMOTE SENSING

Abstract

Large-scale land cover mapping is challenging for its high computational intensity and complexity of data processing pipelines and classification models. Traditional methods are mostly focused on relatively small spatial extents and are limited by sequential computing approaches and related computational models. In order to effectively integrate 3D information from LiDAR with multi-temporal imagery for accurate land cover mapping, an scalable computational framework needs to be developed to process multiple types of geospatial big data with optimal computational workflows. This chapter describes an integrated cyberGIS and deep-learning framework for scalable land cover mapping by seamlessly incorporating a cutting-edge data processing pipeline and classification model with cyberGIS functionalities supported by cyberinfrastructure including the Virtual Roger and Chameleon Cloud services. Specifically, we optimized the process of LiDAR big data processing and analysis using high throughput computing and spatial indexing. The classification model is optimized by adopting a GPU-based SVM algorithm to resolve computational intensity of handling large training datasets using an RBF kernel. We evaluated our framework in a 34-county region in southeast Illinois as a case study and the overall accuracies of 89.33% and 87.29% for original training and class-balanced training scenarios prove the effectiveness of the framework.

5.1 Introduction

Knowledge of spatial distributions of land cover and land use is essential for a variety of scientific fields and societal needs, ranging from environmental and social sciences to natural resource management and urban planning (Chen et al., 2015). Satellite remote sensing has been widely recognized as the most efficient way for mapping land cover and use at different spatial scales and resolutions (Homer et al., 2015). Extensive research has shown the efficiency of using optical imagery (i.e. visible, near and shortwave infrared) for land cover mapping based on 2D features extracted including land surface reflectance, spectral and texture statistics (Almeida et al., 2016; Zhao et al., 2016; Ma et al., 2017). However, land cover and use have complex 3D patterns and optical data can only effectively capture the features distributed horizontally. While a large quantity of large-scale optical data have been collected by Earth orbiting satellites and aircrafts for decades, rapid advancement of airborne Light Detection and Ranging (LiDAR) remote sensing has enabled the collection of large-scale and high-accuracy 3D data opening up a brand new horizon for large-scale mapping of land cover and use from a 3D perspective (Evans et al., 2019). While LiDAR data can be harnessed to obtain rich information about 3D characteristics of various landscape types, 3D feature extraction from LiDAR data is often challenging mainly due to the complexity of such data and computational intensity of related data processing and analysis.

To tackle this challenge, while recent progress on deep learning is promising to achieve high-fidelity 3D feature extraction (Xu et al., 2018), computational research is urgently needed to understand how to enable such approaches to scalable land cover mapping by resolving data and computational intensity (LeCun et al., 2015). Specifically, LiDAR data capture densely distributed 3D points accompanied with sophisticated point attributes (e.g., 3D coordinates,

intensity, return number, scan and incidence angle), which needs tremendous storage space and diverse techniques for data processing and management. Also, the typical large volume and structural complexity of LiDAR data often make related analysis computationally intensive (Papon et al., 2013; Xu et al., 2015; Qi et al., 2017a). Spatial data structures (e.g., octree or kdtree) and related spatial algorithms need to be innovated to achieve desirable computational scalability for handling small footprint and discrete return LiDAR data (Barber et al., 2008; Richter et al., 2013; Xu et al., 2015). How to combine innovative computational and spatial strategies with cutting-edge deep learning methods for 3D feature extraction from LiDAR in the context of large-scale land cover classification is the primary research question of this study.

The recent advance of cyberGIS that is defined as cyber geographic information science and systems based on advanced cyberinfrastructure has enabled solving a variety of data- and computation-intensive geospatial problems across a number of fields including for example agriculture, bioenergy, emergency management, and public health (Wang, 2010; Padmanabhan et al., 2014; Wang, 2016; Hu et al., 2017; Yin et al., 2019; Zhang et al., 2019).

CyberGIS resolves computational challenges by seamlessly integrating interactive spatial analytics, computationally scalable strategies, and straightforward cyberinfrastructure access. Friendly access to cyberGIS capabilities can be gained to overcome the difficulties of intensive computational workloads that exceed the limit of GIS based on sequential computing approaches (Wang et al., 2016). To harness remote sensing data using machine learning approaches, cyberGIS is advantageous to exploit massive data parallelism inherent in such big data (Evans et al., 2019). For example, parameter tuning of machine learning models for analyzing remote sensing data can be extremely computationally intensive since it involves testing hundreds of thousands of parameter combinations. This process can take a long time (e.g., weeks or months)

if only using a personal computer. However, it can be handled effectively and efficiently by using cyberGIS approaches, where computational tasks are divided and conquered simultaneously using many computing elements on cyberinfrastructure. Therefore, this research employs advanced cyberinfrastructure resources - a cyberGIS supercomputer called Virtual ROGER and a GPU resource from the Chameleon cloud to achieve computationally intensive remote sensing data analytics based on an integrated cyberGIS and deep learning framework.

By taking advantage of cyberGIS, we are able to conduct a large-scale land cover classification of a 34-county area of Illinois. First, we conduct LiDAR data preprocessing by splitting the 34 counties of LiDAR data to 17 computing nodes, where each node gets two-county tiles of LiDAR. The individual processors are then used to run data reprojection and noise removal function in parallel for each tile of LiDAR. In total, we processed 89,314 tiles of LiDAR. Second, based on the sample locations and the extents of all the LiDAR tiles, we use a spatial indexing model for efficient sample points queries. Third, we again use the 17 computing nodes for doing data augmentation and voxelization in parallel. Fourth, we extract the training, validation, testing, and prediction data from the generated 3D intensity & occupancy grids and use the GPU resource from the Chameleon cloud for a 3D CNN model training. Fifth, we conduct hyperparameter tuning for model selection based on the training and validation data, and use the model to extract important LiDAR features. Sixth, we combine the extracted LiDAR features with the multispectral imagery to apply an SVM classifier for the classification of seven land cover types. As the last step, we feed the prediction data to the model and generate the classified land cover map in parallel using the 17 computing nodes.

Our earlier research conducted a case study showing the superiority of 3D deep learning for LiDAR feature extraction over previous methods in land cover mapping (Xu et al., 2018).

This research, as a further step, establishes a scalable cyberGIS and deep learning framework for large-scale land cover mapping with high accuracy. As a result, our framework produces an overall accuracy of 89.33% and a kappa value of 0.8859 with proportional class samples and an overall accuracy of 87.29% and a kappa value of 0.8636 with balanced class samples. The experiment result shows the efficiency and scalability of the framework for resolving intensive data and computation in large scale land cover classification. The major contributions of this research are three-fold: (1) This research develops a scalable cyberGIS and deep learning framework for large scale land cover classification using LiDAR big data and multitemporal optical imagery and explore the representative issue of samples with imbalanced distribution over different classes; (2) the framework utilizes high throughput computing for scalable LiDAR data preprocessing (reprojection, denoising, and normalization), processing (point cloud extraction, rotation, cropping, and voxelization), and spatial indexing for efficient spatial queries with large quantities of LiDAR samples; and (3) the framework speedups the classification and prediction processed by taking advantage of a GPU-based SVM algorithm and a multi-GPU computing respectively.

5.2 Study area and dataset

5.2.1 Study area

Our study area covers 34 counties of southeastern Illinois, USA (Figure 5-1). This area has 39,023.63 km^2. It has a mix of humid continental in the northern part and subtropical climatic types in the southern part. Spring is the wettest with extreme weather including tornadoes and winter storms. Summer is humid with warm temperature sometimes reaching above 100 °F (38 °C). Fall is mild with moderate to heavy rainfall. Winter has periodic snow and

it ranges from 0.97m to 0.36m from north to south respectively. In terms of land cover types, agricultural land dominates the majority of the area and concentrates in the northern part. The rest of them are forest, herbaceous, developed, shrubland, wetlands, and water, which are heterogeneously distributed in the middle and southern parts of the study area.

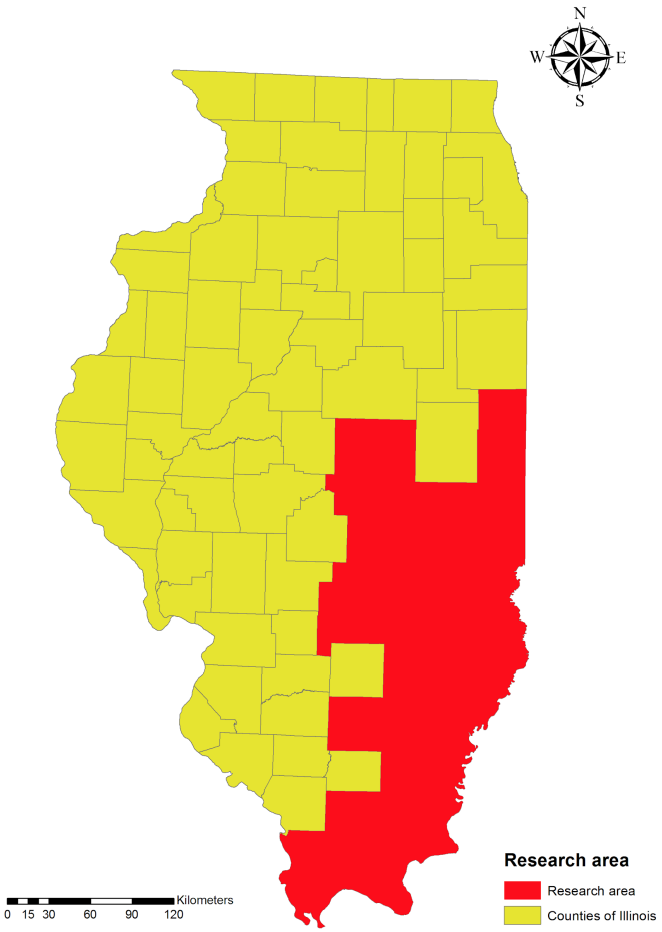


Figure 5-1. Research area

5.2.2 Dataset

89,314 tiles of LiDAR data were flown from 2010 to 2012 covering the research region and downloaded from the Illinois Geospatial Clearing House (https://clearinghouse.isgs.illinois.edu). The data were collected by the Federal Emergency

Management Agency and rectified by multiple companies to ensure both vertical and horizontal accuracy. The average vertical accuracy is 8.53 cm RMSE and average horizontal accuracy is 0.2 meter with a point spacing of 1.4 meters. The instrument collected up to four discrete returns per pulse, with intensity readings of 12-bit dynamic range per measurement, at around 1045nm. The delivered data had an average density of 2 points per m^2 and ranged from 1-5 points per m^2 among different tiles.

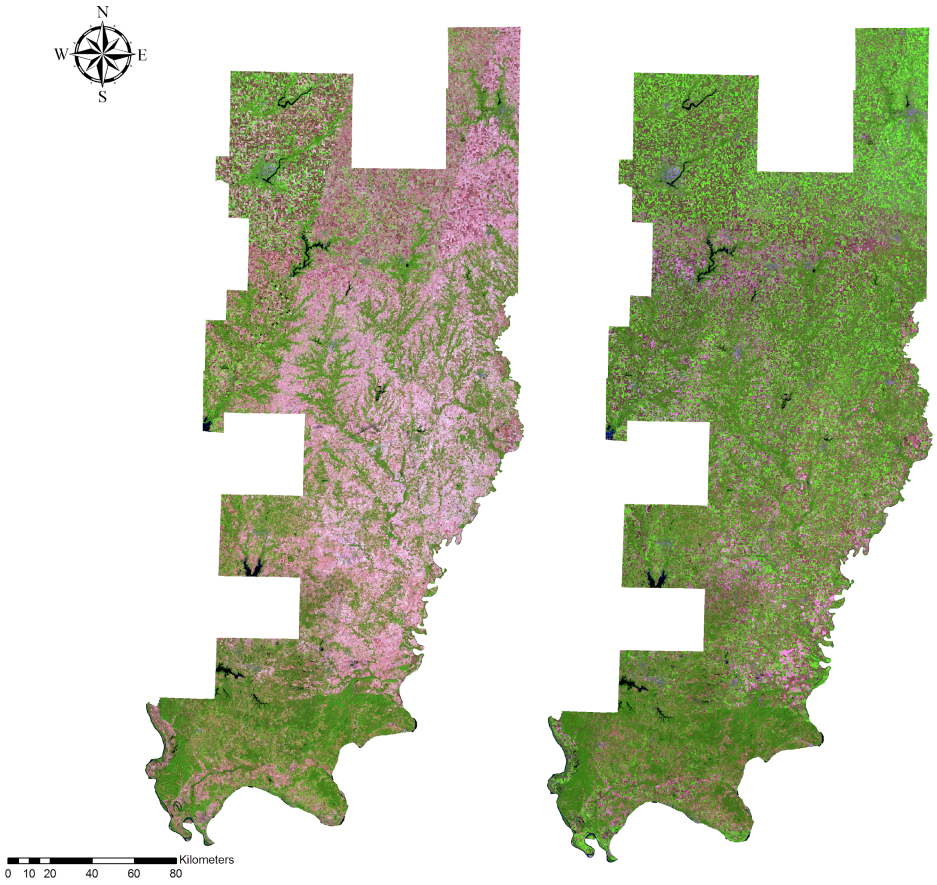
The National Agriculture Imagery Program (NAIP) that was flown in 2010 - 2011 with RGB and NIR bands at a resolution of 1 meter, was downloaded from the USGS Earthexplorer data portal (https://earthexplorer.usgs.gov/) and used as the source to generate reference data through visual interpretation.

This study also collected Landsat 5 images of the study area without cloud and ice coverage, and successfully mosaiced two images shown in Figure 5-2. These two mosaiced images include Thematic Mapper (TM) scenes acquired on August 20, August 27, September 5, September 12, September 28, October 7, October 14 in 2010 with path number from 22 to 23 and row number from 32 to 34. Because all of the images and LiDAR were taken within a two-year period, the land cover was assumed to remain unchanged and thus the same reference data could be used for classification of the dataset. Six bands (blue, green, red, near infrared (IR), and two short wave IR) from each TM image were utilized to provide the spectral information for the classifications. In addition, two indices were calculated – the normalized difference vegetation index (NDVI) (Carlson and Ripley, 1997), and the normalized difference water index (NDWI) (Gao, 1996) – using the following equations:

$$NDVI = \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}} \tag{5.1}$$

$$NDWI = \frac{R_{RED} - R_{SWIR}}{R_{RED} + R_{SWIR}} \tag{5.2}$$

where R_{RED} is the reflectance of the red band, R_{NIR} is the near IR band reflectance, and R_{SWIR} is the shortwave infrared band 5 (1.55–1.75 μm) reflectance of the TM imagery. In addition to spectral content, texture information has long been found effective in the classification process (Haralick et al., 1973). In this study, we calculated five different texture indices including contrast, dissimilarity, homogeneity, energy, and correlation. To summarize, all classifications used the six spectral bands, two indices (NDVI, NDWI), and five texture inputs derived from the spectral information within a 7 × 7 pixel window for a total of 13 inputs per image.



(a) (b)

Figure 5-2. Multitemporal mosaiced Landsat images covering the research area

Mosaiced Landsat imagery 1; (b) Mosaiced Landsat imagery 2.

5.3 Methods

5.3.1 Overview

The framework developed in this research synergistically integrates cyberGIS, cyberinfrastructure, and deep learning. As shown in Figure 5-3, cyberGIS provides application-driven and user-centric functionalities of deep learning through virtual environments built through Anaconda and powered by GPU-based high-performance and cloud computing. It also provides functionalities of data processing with various geospatial toolsets based on cyberGIS software, geospatial big data management through a Linux's EXT4 file system, and a computer cluster with hundreds of CPUs for high throughput computing. All of these functionalities are integrated into a user friendly environment supported by cutting-edge cyberinfrastructure capabilities including Virtual ROGER and Chameleon Cloud.

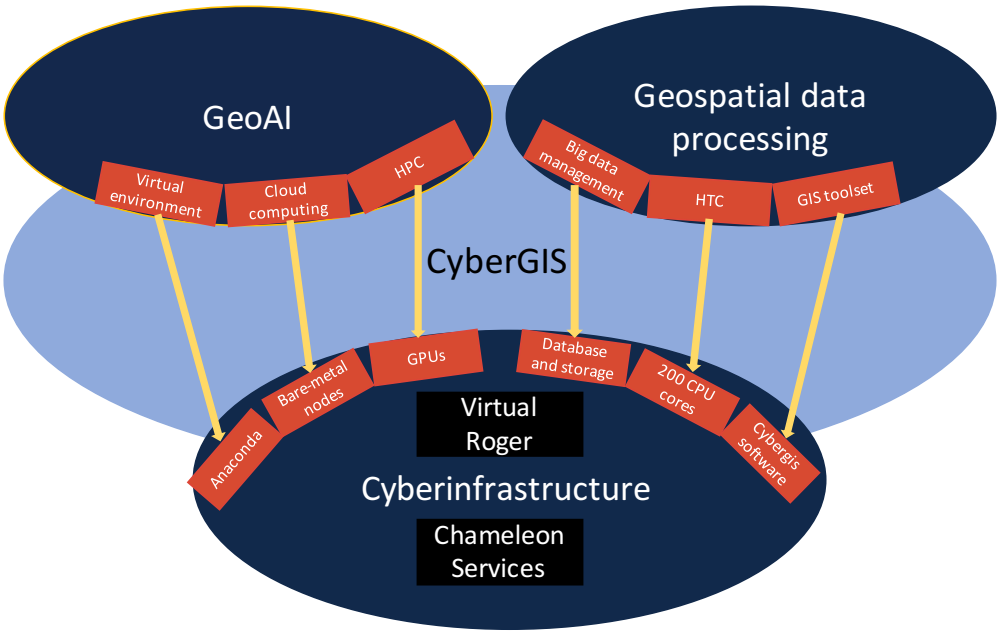


Figure 5-3. The cyberGIS-enabled computational framework

5.3.2 Sampling design

In order to collect representative samples of each class, it is important to keep the inclusion probability the same for each pixel (Xu et al., 2017). Thus, we selected 95,000 pixels randomly over the study area overlaid on the orthophotography. On-screen interpretation of both LiDAR and photography and digitization are utilized to generate reference data. Finally, 93,250 of them are useful and 1,750 are abandoned because of the errors in the aerial photography. The general interpretation principle is that a sample is assigned a class label if that class has a dominant coverage (>50%). One exception is that any sample contains more than 30% of developed type would be classified as developed to consider the connectivity of roads. Explicitly, we define herbaceous as dominant vegetation coverage lower than 1 meter; shrub as dominant vegetation coverage higher than or equal to 1 meter and lower than 5 meters; forest as dominant vegetation coverage higher than or equal to 5 meters. Wetlands are defined as any vegetated area that is saturated with water. Among the 93,250 samples, 60% (55,949) are adopted as training data, 10% (9,325) are used as validation for parameter tuning of deep learning models and 30% (27,976) are adopted as testing for accuracy evaluation. As is shown in Table 5-1, the sample size among different classes are highly imbalanced, in order to explore the influence of imbalanced training samples among different classes, we create two classification scenarios, where the first one utilizes the original imbalanced training samples and the other one utilizes an upsampled version where all the classes are upsampled to have the same number of training data after sample rotation. The validation and testing data sizes remain the same.

Table 5-1. Distribution of training, validation, and testing data among different land cover types

	Agriculture	Developed	Forest	Herbaceous	Shrub	Water	Wetlands	Total
Training	34658	2852	10186	3480	2473	966	1334	55949
Validation	5777	475	1698	580	412	161	222	9325
Testing	17330	1426	5093	1740	1237	483	667	27976

5.3.3 Data preprocessing

Additionally, five texture-based statistics are calculated using grey level co-occurrence matrices (GLCMs) of each image. These layers are stacked as the spectral data for classification. The texture features include contrast, dissimilarity, homogeneity, energy and correlation (Chen et al., 1998). Their formulas are shown in 2.2.3 of this dissertation.

Point data abstraction library (Pdal), a C++ based open source point cloud processing library, is used for LiDAR data preprocessing. LiDAR data are initially reprojected to NAD83 UTM 16N to keep it consistent with spectral data. Statistics-based outlier filter was used to get rid of isolated noisy points including birds, powerlines, etc. We first calculated mean distance of each point to its twelve neighbors. Any point falls below or above two standard deviations of the mean distance of all the points in 3D space would be treated as an outlier and removed. The pipeline of LiDAR data preprocessing is shown in Figure 5-4 below.

5.3.4 Computational framework

Our framework consists of data preprocessing, model training, and testing parts (Figure 5-4). We have a computational bar (red to blue) attached to each operation to indicate the level of computational intensity of each step. In the data preprocessing part, data reprojection has a time complexity of O(n) by constant number of calculations for each point. The noise removal is more computationally challenging for conducting neighborhood searching and has a time complexity of O(nlog(n)). In both of these processes, we use a high-performance computing cluster to run the jobs in parallel on 204 cores. For the spatial queries of samples, we construct a two-level R-tree structure based on the bounding boxes of the counties and their LiDAR tiles for efficient extraction of 78,000,000 samples from the study area. This strategy reduces the time complexity

of this operation from O(n) to O(log(n)). The extracted LiDAR samples are then mapped to the same computer cluster for doing 9 rotations (40 degrees each) to each sample along their z-axis (data augmentation) and voxelization (intensity and occupancy grid creation) in parallel. We did not include agriculture class for sample augmentation since it takes up the majority (62%) of training samples. This process has a time complexity of O(n). In terms of the optical data, we calculated five texture statistics for each image and include them with the spectral information for the final classification and this process has a time complexity of O(n). In model training and testing, we utilize a Tesla M40 GPU for the 3D CNN model training. However, since we have a large number of training samples, the training of the kernel-based SVM classifier becomes computationally challenging because the time complexity ranges between $O(n^2)$ to $O(n^3)$. In our model, we adopted a GPU-based SVM library, ThunderSVM (Wen et al., 2018) to speedup the training and efficiently reduce the training time by 10 times. As the last step, we applied the well-tuned model to the prediction data to generate the final classified map in a voting process by using the same computer cluster for high-throughput computing.

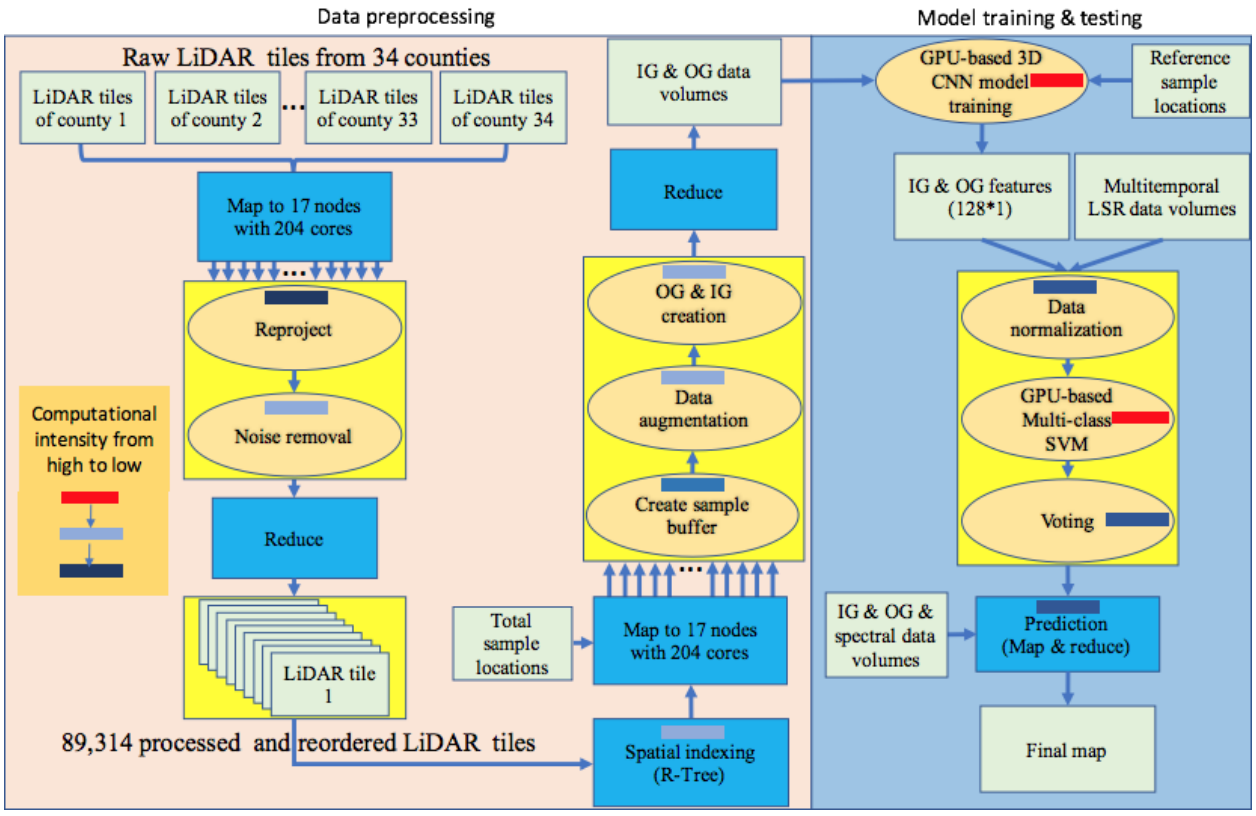


Figure 5-4. The scalable workflow of land cover classification

5.4 Results

5.4.1 Overall performance

Figure 5-5 shows the overall accuracies and kappa values of the two scenarios. The use of original training data provides an overall accuracy and kappa value of 89.33% and 0.8859, while the use of balanced training data yields a 2.05% and 0.0223 lower overall accuracy and kappa respectively. Specifically, the producer's accuracy from the original training data are 75.23%, 72.89%, 75.29%, 85.16%, 69.94%, 71.67%, and 88.96% in agriculture, developed, forest, herbaceous, shrub, water, and wetlands classes and the balanced training data produced higher accuracies in shrub (6.14%) and wetlands (1.8%), and lower accuracies in agriculture (1.95%), developed (0.28%), forest (4.97%), herbaceous (3.39%), and water (1.45%) (Figure 5-6). For the

user's accuracy, the two scenarios produce the same accuracies in agriculture (75.23%), developed (72.89%), forest (75.29%), herbaceous (85.16%), shrub (69.94%), water (71.67%), and wetlands (88.96%) (Figure 5-7).

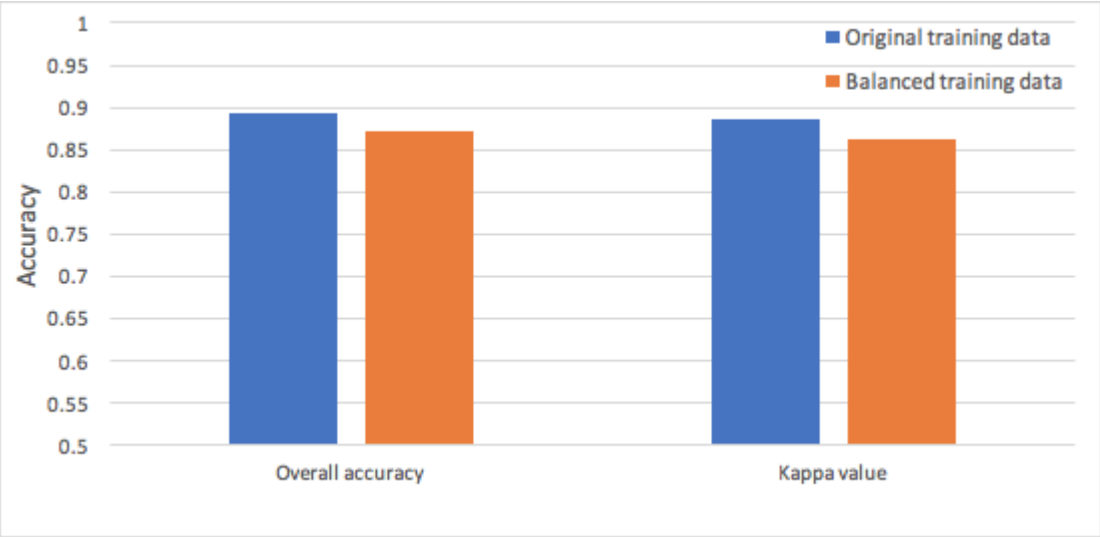


Figure 5-5. Overall accuracies and kappa values from original training data and balanced training data

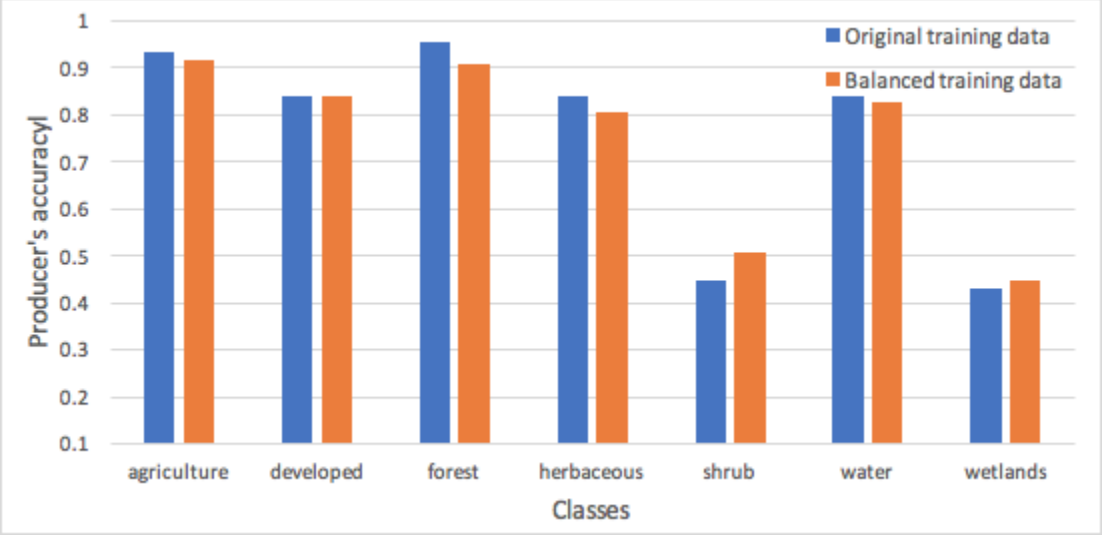


Figure 5-6. Producer's accuracy of original training and balanced training scenarios

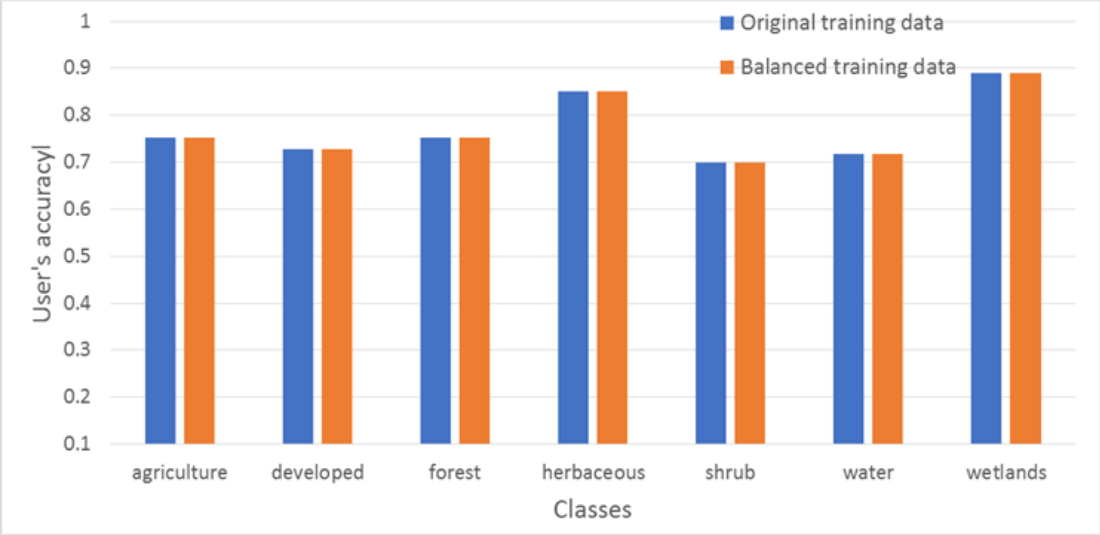


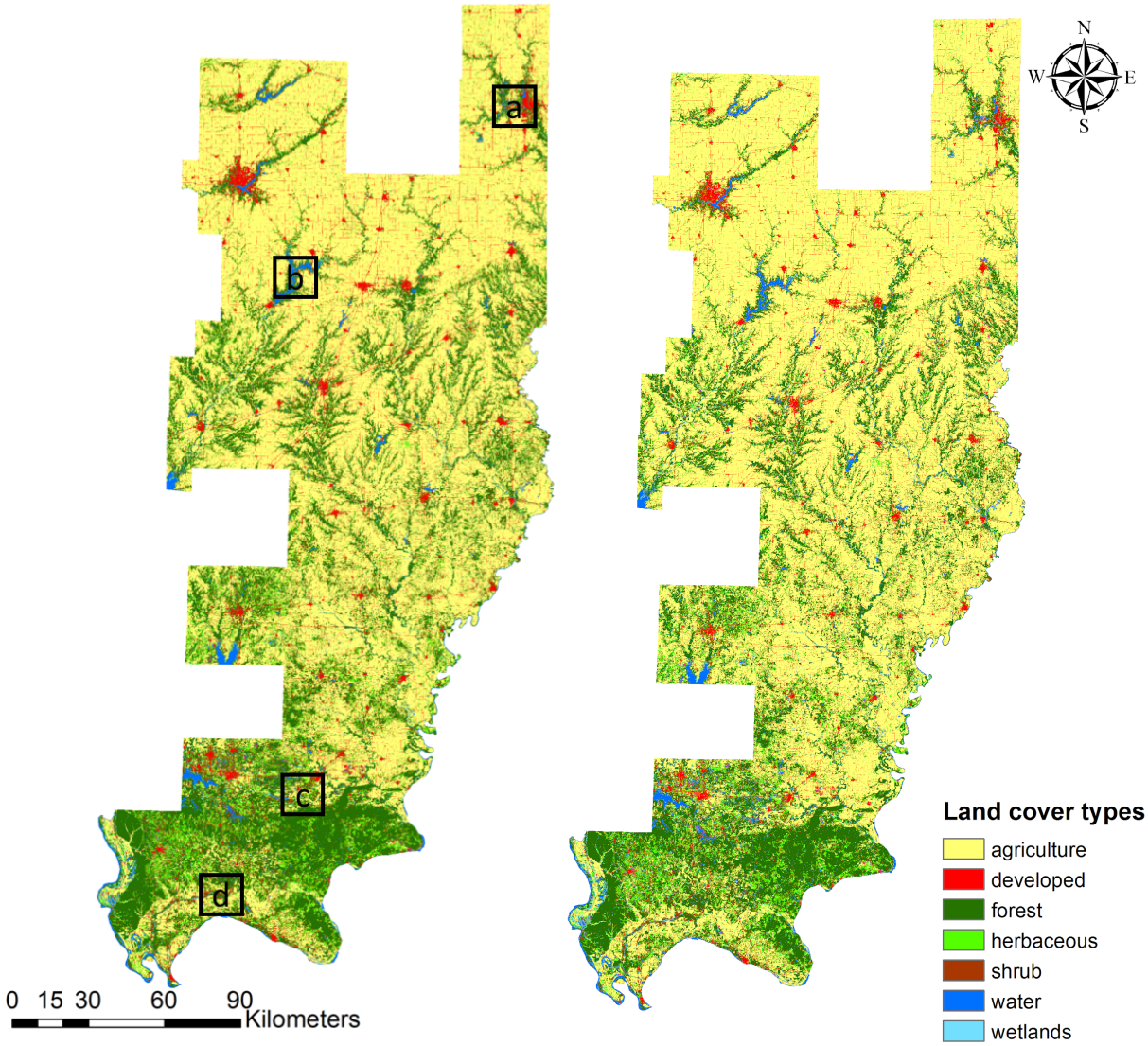
Figure 5-7. User's accuracy of original training and balanced training scenarios

5.4.2 Characteristics of spatial distributions of classified land cover types

The final maps generated from both of the original training and balanced training scenarios are shown in Figure 5-8. Since the two scenarios show a similar pattern, we only discuss the original training scenario which has a higher overall accuracy. We zoom-in four typical locations (Figure 5-9) from north to south to show the details of the classification result.

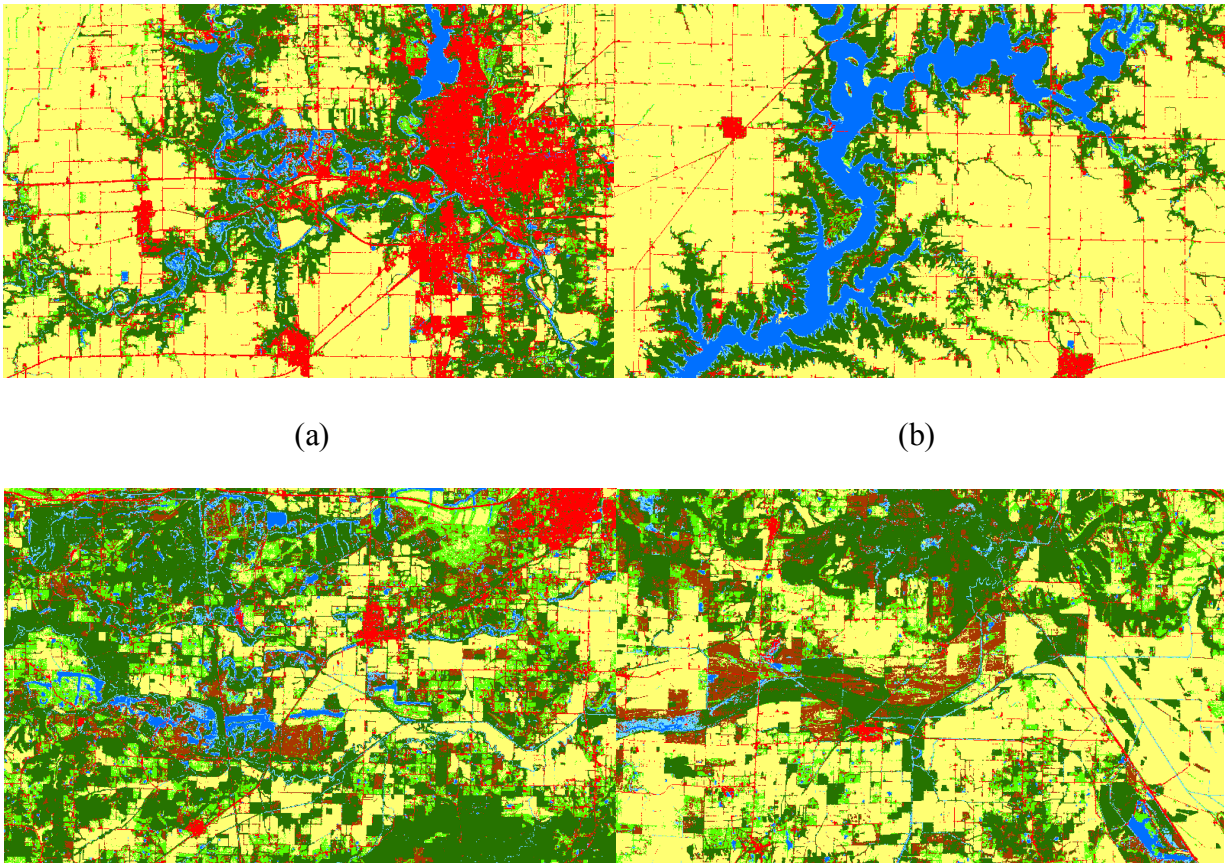
Figure 5-9 (a) shows Danville city of Vermilion county. This area contains highly developed land of downtown Danville, a mix of forest, lakes, and wetlands to the west of the city, and corn fields in the rural area. We can clearly see the spatial pattern of the dense developed area in red and also the extracted road network. The map also reveals the spatial pattern of the natural area to the west of the city, where the forest and wetlands are flourished along the river branches. Figure 5-9 (b) shows the Lake Shelbyvillle located at Shelby county. The boundaries of the lake are accurately extracted and the water pixels enclosed by the boundary are classified smoothly. This area has little dense developed land and is mostly dominated by agricultural land except for the two small towns, one is in the upper left called

Findlay and the other one in the lower right is called Windsor. Most of the roads are country road and they are extracted not as good as the wider roads near cities making some of them disconnected in the rural area. Figure 5-9 (c) shows is natural conservative of Saline county, where the only city, Harrisburg city is located at the upper right corner. This area is rich in surface water and there are many wetlands, riparian herbaceous and shrub growing close to the rivers and lakes. However, the mix of forest and wetlands makes it difficult to distinguish the two. Therefore, the map shows an underestimation of wetlands in the upper left area where most of the pixels are classified as forest class. Figure 5-9 (d) is located at the south boundary of Heron Pond Preserve area at the intersection of Pulaski and Johnson county. This area shows the transition of land cover from forest and wetlands (north) to agricultural land (south). The map shows that shrub and herbaceous are flourished in this transition area, especially along the water area.



(a) (b)

Figure 5-8. Classified land cover map using original and balanced reference data (a: result generated using original reference data; b: result generated using balanced reference data; 1,2,3,4 are the four zoom-in locations shown in Figure 5-9)



(a) (b)

(c) (d)

Figure 5-9. Zoom-in locations of the original training scenario

5.5 Conclusions and Discussion

This chapter describes a scalable framework for land cover classification based on integration of cyberGIS and deep learning using LiDAR data and multi-temporal images. Specifically, this framework utilizes high-throughput computing for scalable LiDAR data processing and spatial indexing for efficient spatial queries with massive LiDAR samples. By utilizing multiple GPUs, the framework further speeds up the classification process 10 times compared to using CPUs. We use reference data generated from both orthophotography and LiDAR for model training, validation and testing. Two scenarios with different training

strategies are constructed to evaluate the framework. In general, we reach an overall accuracy of 89.33% and a kappa value of 0.8859 for the original training data and an overall accuracy of 87.29% and a kappa value of 0.8636 for the balanced training data respectively.

From the computational perspective, the framework seamlessly integrates machine learning and deep learning methods, GPU and HTC computing, and big data management from a user-friendly virtual environment through cyberGIS and supported by cutting-edge cyberinfrastructure capabilities (Virtual ROGER and Chameleon Cloud). From the classification perspective, we compare the results generated from the original training data adopting a random selection strategy and also the balanced training data where training samples from all the classes are upsampled to the same number. Compared to the original training sample scenario, the balanced training sample scenario would compromise on the overall accuracy by 2.04%, but achieve increases of producer's accuracy from the minor classes of wetlands and shrub by 6.14% and 1.8%, respectively. However, the performance on shrub and wetlands classes is still not ideal compared to other classes, which is reflected by the lower producer's accuracy where omission errors dominate.

Our current implementation has the following limitations. The representativeness of the minor classes (shrub and wetlands) is limited. This limitation is reflected from the error matrixes of Table 5-2 and 5-3 that our method underestimated a number of wetlands and shrub samples by wrongly classified them as forest samples. Thus, larger training sizes of the shrub and wetlands classes are necessary in order to ensure the representativeness and expressivity of these two classes in the machine learning and deep learning models. Since the reference data are generated from a large spatial scale based on the aerial photography and LiDAR data collected from different time periods between 2010 to 2012, the variation of seasonality, flight angles, and

different times of a day would add uncertainties to the visual interpretation process and leads to errors in the reference data. For example, when interpreting wetlands samples, we find it sometimes difficult to distinguish it from the forest samples, where the two have similar morphology and spectral reflectance and they are often mingled together.

This research has achieved a novel framework for solving the complex computation, data fusion, and machine learning/deep learning challenges, and produce high quality land cover maps at a large spatial scale through the integration of cyberGIS, cyberinfrastructure, and advanced machine learning and deep learning techniques. Our framework also reveals the performance of the scale-up model on an imbalanced dataset, and potential strategy for improving the accuracy of the underrepresented classes. Future work will focus on exploring effective sampling methods to further enhance the classification accuracy of underrepresented classes, testing the model on more complex landscapes, and potentially applying it to the continental level.

5.6 Supplementary

Table 5-2. Accuracy assessment of classification with original training data

	agriculture	developed	forest	herbaceous	shrub	water	wetlands	User's accuracy
agriculture	16209	29	4	41	7	0	12	0.7523
developed	158	1201	19	94	48	9	39	0.7289
forest	35	31	4871	14	387	6	170	0.7529
herbaceous	822	108	21	1465	211	9	79	0.8516
shrub	84	42	123	109	553	3	48	0.6994
water	6	2	2	1	2	406	32	0.7167
wetlands	16	13	53	16	29	50	287	0.8896
Producer's accuracy	0.9353	0.8422	0.9564	0.8420	0.4470	0.8406	0.4303	
Overall accuracy	0.8933	Kappa	0.8859					

Table 5-3. Accuracy assessment of classification with balanced training data

	agriculture	developed	forest	herbaceous	shrub	water	wetlands	User's accuracy
agriculture	15871	31	6	42	8	2	8	0.7523
developed	228	1197	29	98	57	6	32	0.7289
forest	114	31	4618	11	283	4	147	0.7529
herbaceous	947	97	14	1406	214	16	80	0.8516
shrub	106	50	308	153	629	2	65	0.6994
water	15	4	5	5	2	399	36	0.7167
wetlands	49	16	113	25	44	54	299	0.8896
Producer's accuracy	0.9158	0.8394	0.9067	0.8080	0.5085	0.8261	0.4483	
Overall accuracy	0.8729	Kappa	0.8636					

CHAPTER 6: CONCLUDING DISCUSSION

6.1 Summary of contributions

In this dissertation, we demonstrate four a cyberGIS-enabled deep learning framework for remote sensing data analytics for mapping landscape patterns and dynamics (Figure 6-1) through the following four interrelated studies.

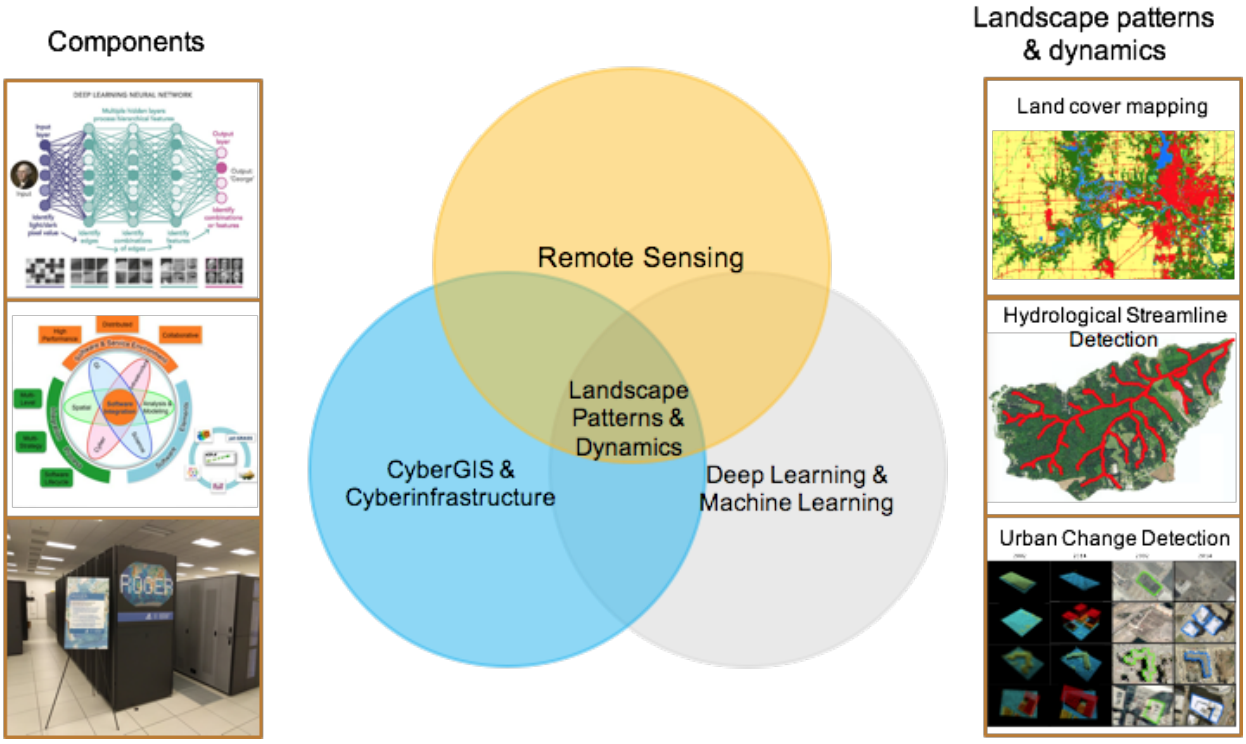


Figure 6-1. CyberGIS-enabled landscape patterns and dynamics studies

6.1.1 Land cover classification

The classification results show that our method outperforms the traditional method by 4.23% (from 78.87% to 83.71%) when solely using LiDAR and 1.84% (from 90.54% to 92.37%) when combining all available images. We demonstrate that our method can effectively extract LiDAR features and improve fine-scale land cover mapping through fusion of complementary

types of remote sensing data. To be specific, we find that the features derived from our method are promising for capturing extensive complex features of various land cover types in three dimensions that traditional methods are not designed to resolve. For example, the benchmark method fails to classify many wetlands pixels located along rivers or lakes in a dense forest, developed regions that are shaded by surrounding vegetation, shrub and herbaceous pixels that share similar spectral information and growing patterns and turbid water pixels, but they are correctly classified by our method. Our validation process using orthophotography and LiDAR further proves that the new method can achieve high-accuracy land cover mapping.

6.1.2 Urban building change detection

The high classification accuracy proves the effectiveness of the method for the classification of various types of urban buildings at individual building level even with the datasets of limited point accuracy, density, and a low horizontal and vertical accuracy. From the change detection perspective, our method avoids detecting changes of non-building objects, wrong or incomplete detection of building changes and achieves a high level of completeness and correctness, and reliable and accurate estimation of different types of urban changes. In previous studies, point cloud registration is often the first step for conducting change detection. In our method, we eliminate this computationally intensive process by directly working on individual buildings, where the footprints of individual buildings are delineated and overlapping ones are formed into pairs for comparison of changes. Our method conducts change detection by estimating the change of building volume and footprint, which provides a new way for quantitatively estimating the amount of building change volume at the level of individual buildings.

6.1.3 Hydrological streamline detection

In this research, we proposed to use an attention U-net model for hydrological streamline detection using eight different LiDAR-derived feature maps. The result shows that our method outperforms traditional machine learning models and flow accumulation methods by providing a smoother and better connectivity of streamline and water body features. In order to test the model thoroughly, we create four different scenarios by splitting our research area into upper/lower and left/right portions for generating training/validation and testing patches respectively. The attention U-net model used in our research generates F1-scores of 83.02%, 90.53%, 91.91%, and 80.79% across the four scenarios, which outperforms the best benchmark by 8.61%, 9.39%, 13.68%, and 13.31%.

6.1.4 Scaling up of land cover classification

This research develops a scalable land cover classification method for solving the complex computation, data fusion, and deep learning challenges, and produce high quality land cover maps at a large spatial scale through the integration of cyberGIS, cyberinfrastructure, and advanced machine learning and deep learning techniques. Our method also reveals the performance of the scale-up model on an imbalanced dataset, and potential strategy for improving the accuracy of the underrepresented classes. In general, we reach an overall accuracy of 89.33% and a kappa value of 0.8859 for the original training data and an overall accuracy of 87.29% and a kappa value of 0.8636 for the balanced training data respectively.

6.1.5 Summary

The major contributions of these studies are highlighted as follows. The first study delves into land cover classification by developing a deep learning framework for the extraction of LiDAR features and fusing with multitemporal optical imagery to generate a high accuracy a land cover map at the county level. The major contributions of this research are two-fold: (1) a novel method for land cover feature extraction by using full geometric and intensity information from LiDAR data, and (2) an effective data fusion model for using 3D features from LiDAR and multi-temporal imagery to achieve high accuracy in land cover classification.

The second study centers on using 3D deep learning and a rule-based method to extract individual buildings, detect different types of changes, and estimate the amount of changes. The central contributions of this research are summarized as follows: (1) This research adapts the cutting-edge PointSIFT deep learning algorithm (PointSIFT) for point labeling to airborne LiDAR classification by integrating the iterative farthest point algorithm for training sample standardization, a multi-patch-based classification strategy, and a point-based voting based on different sizes of training patches to achieve high accuracy of classification. (2) Our research conducts building change detection by estimating the change of building volumes and footprints. It provides a new way for quantitatively estimating urban changes in 3D at a fine spatial scale. (3) Our method eliminates the computation-intensive point cloud registration in previous studies, by directly working on individual buildings, where the footprints of individual buildings are delineated and overlapping ones are formed into pairs for comparison of changes. This design avoids the difficulty of distinguishing between non-building and building changes, which many previous studies struggled with. (4) Although high quality (+20 pts/m2, <0.1 RMSE) multitemporal LiDAR data have been proven to be effective for 3D urban modeling and change

detection, this research demonstrates that it is still possible to get high accuracy results by using readily available LiDAR data (3 pts/m2) with a large error rate (0.5 RMSE).

The third study focuses on generating high accuracy streamline maps using deep learning. In this research, we construct a fully convolutional neural network model with attention module for hydrological feature extraction and utilize the feature concatenation structure in the U-net model to generate high accuracy streamline maps. The contributions of our research are two-fold: (1) This research constructs a novel model for automatic hydrological streamline detection using deep learning; (2) The model outperforms both traditional machine learning methods and two hydrological flow accumulation models by generating topologically smooth and connected results and overcomes the overestimation over dried drainage lines of low accumulation models.

The fourth study develops a computationally scalable model for large-scale land cover classification over 34 counties of southeast Illinois. The major contributions of this research are three-fold: (1) This research is the first to develop a scalable deep learning method for large scale land-cover classification using LiDAR big data and multitemporal optical imagery and explore the representative issue of samples with imbalanced distribution over different classes; (2) This research utilizes high-throughput computing for scalable LiDAR data preprocessing (reprojection, denoising, and normalization), processing (point cloud extraction, rotation, cropping, and voxelization), and spatial indexing for efficient spatial queries with large quantities of LiDAR samples; and (3) This research speeds up the classification and prediction processes by adopting a GPU-based SVM algorithm and a multi-GPU approach.

6.2 Limitations and future work

In the first study in terms of land cover classification, errors in the reference imagery and LiDAR data limit the performance of our model for the land cover classification research. From the imagery perspective, errors are introduced by clouds and shadows in the imagery due to various acquisition times and dates. Errors in LiDAR data come from the missing data of some places caused by occlusion from physical obstacles or large flight angles. In these cases, real features of underneath objects cannot be learned and correctly extracted, which in turn could cause errors in the results. In areas where vegetation or other coverage is dense, ground returns may also miss from detection and raise errors. Furthermore, misalignment between LiDAR and Landsat imagery could also cause errors. Our future work will focus on exploring the optimal spatial resolution of voxels, understanding the deep features extracted and testing areas with more complex landscapes.

In the second study of urban building extraction and change detection, errors of our model come from two primary aspects: the classification and change detection process. In the process of classification, our method misses a number of temporarily constructed prefabricated houses that are similar (size, shape, and color) to the large containers (non-building) in the residential and industrial areas. The method also fails to recognize some irregularly shaped buildings in the commercial area. Because of the abundant noise in the 2002 data, the algorithm failed to isolate a few adjacent buildings in the residential area. From the perspective of change detection, our method makes some false change detections on several small buildings near high-rise ones, where their points are partially blocked and vary due to different flight angles. Also, some non-building objects closely attached to buildings that change over time are wrongly recognized as changes. The third aspect comes from the data quality. The data collection

standards and flight parameters have a large variation between 2002 and 2014, where the 2002 LiDAR dataset has a much lower accuracy and higher noise rate compared to the 2014 one. This discrepancy leads to part of the false cases. In the future, we need to further refine the method to make it more resilient to the aforementioned error sources. Specifically, training data need to be improved by including additional building types, such as prefabricated houses and irregularly shaped buildings. For change detection, we plan to study pairwise individual building registration in order to extract more details of the changed buildings. Furthermore, we plan to test our method over other study sites and various qualities of datasets for sensitivity analysis to verify its stability and generality.

In the third study of hydrological streamline detection, our experiments of streamline mapping are conducted using the reference data generated from ground surveying of head waters and visual delineation of streamline channels based on the LiDAR feature maps. We utilize a fixed buffer size for streamline channels - 3m in order to include enough information for the streamline class, which would introduce pixels from non-stream classes as errors of the narrower stream channels to the stream class. Considering the width of streamlines would change over time during a year, so in this study we focus on more of the topological result instead of the exact width of stream channels by using a compensated accuracy evaluation method over our model and benchmark methods. In the future, we need to focus on extracting more details and classify different types of stream channels using multi-temporal datasets. On the other hand, our model is applied at a study area with high quality LiDAR available – 43 pts/m2 at county level, which would need to be further verified at other locations and applied to a larger spatial scale in the future.

As the fourth study of the scaleup of land cover classification, we find that the performance of our model on minor classes (shrub and wetlands) is not as good because of limited number of reference samples. In the future, we need to further refine our sample generation strategy by specifically generating more samples for the minor classes (shrub and wetlands) to ensure representation in the deep learning model. Moreover, since it is generated from a large spatial scale based on the aerial photography and LiDAR data collected from different time periods between 2010 to 2012, the variation of seasonality, flight angles, and different times of a day would add uncertainties to the visual interpretation process and leads to errors in the reference data. For example, when interpreting wetlands samples, we find it sometimes difficult to distinguish from the forest samples, where the two have similar morphology and spectral reflectance and they are often mingled together. In the future, we need to combine multiple reliable reference sources including ground surveying or GIS databases to refine our reference dataset. We also need to test the model on more complex landscapes, and potentially applying it to the continental level.

REFERENCES

[1]. Agarap, A. F. (2018). Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*.

[2]. Aghsaei, H., Dinan, N. M., Moridi, A., Asadolahi, Z., Delavar, M., Fohrer, N., & Wagner, P. D. (2020). Effects of dynamic land use/land cover change on water resources and sediment yield in the Anzali wetland catchment, Gilan, Iran. *Science of The Total Environment*, *712*, 136449.

[3]. Albert, A., Kaur, J., & Gonzalez, M. C. (2017, August). Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1357-1366).

[4]. Alexandre, L. A. (2016). 3D object recognition using convolutional neural networks with transfer learning between input channels. In *Intelligent Autonomous Systems 13* (pp. 889-898). Springer, Cham.

[5]. Almeida, C. A. D., Coutinho, A. C., Esquerdo, J. C. D. M., Adami, M., Venturieri, A., Diniz, C. G., Dessay, N., Durieux, L., & Gomes, A. R. (2016). High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. *Acta Amazonica*, *46*(3), 291-302.

[6]. Alonzo, M., Bookhagen, B., McFadden, J. P., Sun, A., & Roberts, D. A. (2015). Mapping urban forest leaf area index with airborne lidar using penetration metrics and allometry. *Remote Sensing of Environment*, *162*, 141-153.

[7]. Alvioli, M., Mondini, A. C., Fiorucci, F., Cardinali, M., & Marchesini, I. (2018). Topography-driven satellite imagery analysis for landslide mapping. *Geomatics, Natural Hazards and Risk*, *9*(1), 544-567.

[8]. Anderson, D. L. (2012). *Detailed hydrographic feature extraction from high-resolution LIDAR data* (No. INL/EXT-12-26969). Idaho National Laboratory (INL)

[9]. Anderson, J. R. (1976). *A land use and land cover classification system for use with remote sensor data* (Vol. 964). US Government Printing Office.

[10]. Antonarakis, A. S., Richards, K. S., & Brasington, J. (2008). Object-based land cover classification using airborne LiDAR. *Remote Sensing of environment*, *112*(6), 2988-2998.

[11]. Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., & Savarese, S. (2016). 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1534-1543).

[12]. Arroyo, L. A., Johansen, K., Armston, J., & Phinn, S. (2010). Integration of LiDAR and QuickBird imagery for mapping riparian biophysical parameters and land cover types in Australian tropical savannas. *Forest Ecology and Management*, *259*(3), 598-606.

[13]. Asra, G. (1989). *Theory and applications of optical remote sensing*. G. Asrar (Ed.). New York: Wiley.

[14]. Audebert, N., Le Saux, B., & Lefèvrey, S. (2017, March). Fusion of heterogeneous data in convolutional networks for urban semantic labeling. In *2017 Joint Urban Remote Sensing Event (JURSE)* (pp. 1-4). IEEE.

[15]. Bagley, J. E., Desai, A. R., Harding, K. J., Snyder, P. K., & Foley, J. A. (2014). Drought and deforestation: Has land cover change influenced recent precipitation extremes in the Amazon?. *Journal of Climate*, *27*(1), 345-361.

[16]. Ban, Y., & Yousif, O. A. (2012). Multitemporal spaceborne SAR data for urban change detection in China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(4), 1087-1094.

[17]. Barber, D. M., Holland, D., & Mills, J. P. (2008). Change detection for topographic mapping using three-dimensional data structures. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *37*(B4), 1177-1182.

[18]. Bartels, M., & Wei, H. (2006, September). Rule-based improvement of maximum likelihood classified LIDAR data fused with co-registered bands. In *Annual Conference of the Remote Sensing and Photogrammetry Society, CD Proceedings* (Vol. 5, pp. 1-9).

[19]. Bennett, M. M., & Smith, L. C. (2017). Advances in using multitemporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics. *Remote Sensing of Environment*, *192*, 176-197.

[20]. Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. *Surveys in Geophysics*, *37*(2), 307-337.

[21]. Boulos, M. N. K., Peng, G., & VoPham, T. (2019). An overview of GeoAI applications in health and healthcare.

[22]. Bouziani, M., Goïta, K., & He, D. C. (2010). Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 143-153.

[23]. Bozkurt, S., Elibol, G., Gunal, S., & Yayan, U. (2015, September). A comparative study on machine learning algorithms for indoor positioning. In *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)* (pp. 1-8). IEEE.

[24]. Brennan, R., & Webster, T. L. (2006). Object-oriented land cover classification of lidar-derived surfaces. *Canadian journal of remote sensing*, *32*(2), 162-172.

[25]. Bretar, F., Chauve, A., Bailly, J. S., Mallet, C., & Jacome, A. (2008). Terrain surfaces and 3D landcover classification from small footprint full-waveform lidar data: Application to badlands.

[26]. Brovkin, V., Boysen, L., Arora, V. K., Boisier, J. P., Cadule, P., Chini, L., Claussen, M., Friedlingstein, P., Gayler, V., Van Den Hurk, B.J.J.M., & Hurtt, G.C. (2013). Effect of anthropogenic land-use and land-cover changes on climate and land carbon storage in CMIP5 projections for the twenty-first century. *Journal of Climate*, *26*(18), 6859-6881.

[27]. Campbell, J. B., & Wynne, R. H. (2011). *Introduction to remote sensing*. Guilford Press.

[28]. Carlberg, M., Gao, P., Chen, G., & Zakhor, A. (2009, November). Classifying urban landscape in aerial LiDAR using 3D shape analysis. In *2009 16th IEEE International Conference on Image Processing (ICIP)* (pp. 1701-1704). IEEE.

[29]. Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, *62*(3), 241-252.

[30]. Caruso, B. S. (2014). GIS-Based Stream Classification in a Mountain Watershed for Jurisdictional Evaluation. *JAWRA Journal of the American Water Resources Association*, *50*(5), 1304-1324.

[31]. Charaniya, A. P., Manduchi, R., & Lodha, S. K. (2004, June). Supervised parametric classification of aerial lidar data. In *2004 Conference on Computer Vision and Pattern Recognition Workshop* (pp. 30-30). IEEE.

[32]. Chen, B., Shi, S., Gong, W., Zhang, Q., Yang, J., Du, L., Sun, J., Zhang, Z., & Song, S. (2017). Multispectral LiDAR point cloud classification: A two-step approach. *Remote Sensing*, *9*(4), 373.

[33]. Chen, C., & Pau, L. (1998). PW (eds.), The Handbook of Pattern Recognition and Computer Vision.

[34]. Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng., S., Lu, M., Zhang, W., Tong, X., & Mills, J. (2015). Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, 7-27.

[35]. Chen, Y., Li, C., Ghamisi, P., Shi, C., & Gu, Y. (2016, July). Deep fusion of hyperspectral and LiDAR data for thematic classification. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3591-3594). IEEE.

[36]. Chen, Y., Su, W., Li, J., & Sun, Z. (2009). Hierarchical object oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research*, *43*(7), 1101-1110.

[37]. Chorley, R. J., & Dale, P. F. (1972). Cartographic problems in stream channel delineation. *Cartography*, *7*(4), 150-162.

[38]. Clement, M. A., Kilsby, C. G., & Moore, P. (2018). Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management*, *11*(2), 152-168.

[39]. Clifton, W. E., Steele, B., Nelson, G., Truscott, A., Itzler, M., & Entwistle, M. (2015, May). Medium altitude airborne Geiger-mode mapping LIDAR system. In *Laser Radar*

Technology and Applications XX; and Atmospheric Propagation XII (Vol. 9465, p. 946506). International Society for Optics and Photonics.

[40]. Clubb, F. J., Bookhagen, B., & Rheinwalt, A. (2019). Clustering river profiles to classify geomorphic domains. *Journal of Geophysical Research: Earth Surface*, *124*(6), 1417-1439.

[41]. Cohen, T. S., Geiger, M., Köhler, J., & Welling, M. (2018). Spherical cnns. *arXiv preprint arXiv:1801.10130*.

[42]. Colson, T. P. (2006). Stream network delineation from high-resolution digital elevation models.

[43]. Colson, T., Gregory, J., Dorney, J., & Russell, P. (2008). Topographic and soil maps do not accurately depict headwater stream networks. *National Wetlands Newsletter*, *30*(3), 25-28.

[44]. Comer, P., Faber-Langendoen, D., Evans, R., Gawler, S., Josse, C., Kittel, G., Menard, S., Pyne, M., Reid, M., Schulz, K., & Snow, K. (2003). Ecological systems of the United States: a working classification of US terrestrial systems. *NatureServe, Arlington, VA*, *75*.

[45]. Costa, M. H., Coe, M. T., & Galbraith, D. R. (2016). Land-atmosphere interactions.

[46]. Coutard, O., & Rutherford, J. (Eds.). (2015). *Beyond the networked city: Infrastructure reconfigurations and urban change in the North and South*. Routledge.

[47]. Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(5), 1416-1427.

[48]. Darom, T., & Keller, Y. (2012). Scale-invariant features for 3-D mesh models. *IEEE Transactions on Image Processing*, *21*(5), 2758-2769.

[49]. De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., Chu, W., Antrop, M., De Maeyer, P., Finke, P., Van Meirvenne, M., Verniers, J., & Crombé, P. (2013). Application of the topographic position index to heterogeneous landscapes. *Geomorphology*, *186*, 39-49.

[50]. Deng, C., Li, C., Zhu, Z., Lin, W., Xi, L. (2017). Evaluating the impacts of atmospheric correction, seasonality, environmental settings, and multi-temporal images on subpixel urban impervious surface area mapping with Landsat data. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2017.09.015.

[51]. Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and trends in signal processing*, *7*(3–4), 197-387.

[52]. Devaney, J., Barrett, B., Barrett, F., Redmond, J., & John, O. (2015). Forest cover estimation in Ireland using radar remote sensing: a comparative analysis of forest cover assessment methodologies. *Plos one*, *10*(8), e0133583.

[53]. Doneus, M. (2013). Openness as visualization technique for interpretative mapping of airborne lidar derived digital terrain models. *Remote sensing*, *5*(12), 6427-6442.

[54]. Donihue, C. M., & Lambert, M. R. (2015). Adaptive evolution in urban ecosystems. *Ambio*, *44*(3), 194-203.

[55]. Du, S., Zhang, Y., Qin, R., Yang, Z., Zou, Z., Tang, Y., & Fan, C. (2016). Building change detection using old aerial images and new LiDAR data. *Remote Sensing*, *8*(12), 1030.

[56]. ESRI 2016. ArcGIS Desktop: Release 10.4. Redlands, CA: Environmental Systems Research Institute.

[57]. Evans, M. R., Oliver, D., Yang, K., Zhou, X., Ali, R. Y., & Shekhar, S. (2019). Enabling spatial big data via CyberGIS: Challenges and opportunities. In *CyberGIS for geospatial discovery and innovation* (pp. 143-170). Springer, Dordrecht.

[58]. Firozjaei, M. K., Kiavarz, M., Alavipanah, S. K., Lakes, T., & Qureshi, S. (2018). Monitoring and forecasting heat island intensity through multi-temporal image analysis and cellular automata-Markov chain modelling: A case of Babol city, Iran. *Ecological Indicators*, *91*, 155-170.

[59]. Flitton, G., Breckon, T. P., & Megherbi, N. (2012, June). A 3D extension to cortex like mechanisms for 3D object class recognition. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3634-3641). IEEE.

[60]. Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., & Alcamo, J. (2013). Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study. *Global Environmental Change*, *23*(1), 144-156.

[61]. Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., & Helkowski, J. H. (2005). Global consequences of land use. *science*, *309*(5734), 570-574.

[62]. Fritz, K. M., Hagenbuch, E., D'Amico, E., Reif, M., Wigington Jr., P.J., Leibowitz, S.G., Comeleo, R.L., Ebersole, J.L., & Nadeau. T.L. (2013). Comparing the extent and permanence of headwater streams from two field surveys to values from hydrographic databases and maps. *JAWRA Journal of the American Water Resources Association, 49*(4), 867–882.

[63]. Furuya, T., & Ohbuchi, R. (2009, July). Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features. In *Proceedings of the ACM international conference on image and video retrieval* (pp. 1-8).

[64]. Gao, B. C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, *58*(3), 257-266.

[65]. García-Gutiérrez, J., Mateos-García, D., Garcia, M., & Riquelme-Santos, J. C. (2015). An evolutionary-weighted majority voting and support vector machines applied to contextual classification of LiDAR and imagery data fusion. *Neurocomputing*, *163*, 17-24.

[66]. Glaeser, E. L. & Ward, B. A. (2009). The causes and consequences of land use regulation: Evidence from Greater Boston. *Journal of Urban Economics, 65*(3), 265-278.

[67]. Golden, C., Murtha, T., Cook, B., Shaffer, D. S., Schroder, W., Hermitt, E. J., Firpi, O. A., & Scherer, A. K. (2016). Reanalyzing environmental lidar data for archaeology: Mesoamerican applications and implications. *Journal of Archaeological Science: Reports*, *9*, 293-308.

[68]. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

[69]. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, *202*, 18-27.

[70]. Guo, J., Liu, H., Li, Z., Rosenfeld, D., Jiang, M., Xu, W., Jiang, J.H., He, J., Chen, D., Min, M., & Zhai, P. (2018). Aerosol-induced changes in the vertical structure of precipitation: a perspective of TRMM precipitation radar. *Atmospheric Chemistry and Physics, 18*(18), 13329-13343.

[71]. Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(1), 56-66.

[72]. Gupta, S., Girshick, R., Arbeláez, P., & Malik, J. (2014, September). Learning rich features from RGB-D images for object detection and segmentation. In *European conference on computer vision* (pp. 345-360). Springer, Cham.

[73]. Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621.

[74]. Harrison, J., & Hoyler, M. (Eds.). (2015). *Megaregions: Globalization s New Urban Form?*. Edward Elgar Publishing.

[75]. Hata, A. Y., & Wolf, D. F. (2015). Feature detection for vehicle localization in urban environments using a multilayer LIDAR. *IEEE Transactions on Intelligent Transportation Systems*, *17*(2), 420-429.

[76]. Heidemann, H.K. (2018). *Lidar base specification* (ver. 1.3, February 2018): U.S. Geological Survey Techniques and Methods, book 11, chap. B4, 101 p., https://doi.org/10.3133/tm11b4.

[77]. Hipps, R., Chopra, T., Zhao, P., Kwartler, E., & Jaume, S. (2017). Geospatial Analytics to Improve the Safety of Autonomous Vehicles. *International Journal of Knowledge-Based Organizations (IJKBO)*, *7*(3), 40-51.

[78]. Hoegberg, S. (2016). *National Hydrography Requirements and Benefits Study - Preliminary Results: Dewberry LLC, Fairfax, Virginia*. 140p. https://www.usgs.gov/core-science-systems/ngp/national-hydrography/hydrography-requirements-and-benefits-study.

[79]. Höft, N., Schulz, H., & Behnke, S. (2014, September). Fast semantic segmentation of RGB-D scenes with GPU-accelerated deep neural networks. In *Joint German/Austrian*

Conference on Artificial Intelligence (Künstliche Intelligenz) (pp. 80-85). Springer, Cham.

[80]. Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*, *81*(5), 345-354.

[81]. Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow A., VanDriel, J.N., & Wickham, J. (2007). Completion of the 2001 national land cover database for the counterminous United States. *Photogrammetric engineering and remote sensing*, *73*(4), 337.

[82]. Hu, Y., Gao, S., Newsam, S., & Lunga, D. (2019a). GeoAI 2018 workshop report the 2nd ACM SIGSPATIAL international workshop on GeoAI: AI for geographic knowledge discovery seattle, WA, USA-November 6, 2018. *ACM SIGSPATIAL special*, *10*(3), 16-16.

[83]. Hu, H., Lin, T., Wang, S., & Rodriguez, L. F. (2017). A cyberGIS approach to uncertainty and sensitivity analysis in biomass supply chain optimization. *Applied energy*, *203*, 26-40.

[84]. Hu, Y., Mao, H., & McKenzie, G. (2019b). A natural language processing and geospatial clustering framework for harvesting local place names from geotagged housing advertisements. *International Journal of Geographical Information Science*, *33*(4), 714-738.

[85]. Huang, M. J., Shyue, S. W., Lee, L. H., & Kao, C. C. (2008). A knowledge-based approach to urban feature classification using aerial imagery with lidar data. *Photogrammetric Engineering & Remote Sensing*, *74*(12), 1473-1485.

[86]. Huang, X., Wen, D., Li, J., & Qin, R. (2017). Multi-level monitoring of subtle urban changes for the megacities of China using high-resolution multi-view satellite imagery. *Remote sensing of environment*, *196*, 56-75.

[87]. Isenburg M. 2011. "LAStools - efficient LiDAR processing software" (version 141017, unlicensed), obtained from http://rapidlasso.com/LAStools

[88]. Isenburg, M., Liu, Y., Shewchuk, J., & Snoeyink, J. (2006). Streaming computation of Delaunay triangulations. In *ACM SIGGRAPH 2006 Papers* (pp. 1049-1056).

[89]. Jenson, S. K., & Domingue, J. O. (1988). Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric engineering and remote sensing*, *54*(11), 1593-1600.

[90]. Ji, S., Xu, W., Yang, M., & Yu, K. (2012). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 221-231.

[91]. Ji, S., Zhang, C., Xu, A., Shi, Y., & Duan, Y. (2018). 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, *10*(1), 75.

[92]. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., & Darrell, T. (2014, November). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 675-678).

[93]. Jiang, M., Wu, Y., Zhao, T., Zhao, Z., & Lu, C. (2018). Pointsift: A sift-like network module for 3d point cloud semantic segmentation. *arXiv preprint arXiv:1807.00652*.

[94]. Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis*, *36*, 61-78.

[95]. Kampffmeyer, M., Salberg, A. B., & Jenssen, R. (2016). Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 1-9).

[96]. Kim, M., Lee, J., Han, D., Shin, M., Im, J., Lee, J., Han, D., Shin, M., Im, J., Lee, J., Quackenbush, L.J., & Gu, Z. (2018). Convolutional neural network-based land cover classification using 2-D spectral reflectance curve graphs with multitemporal satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(12), 4604-4617.

[97]. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[98]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

[99]. Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, *14*(5), 778-782.

[100]. Larkin, A., & Hystad, P. (2019). Evaluating street view exposure measures of visible green space for health research. *Journal of exposure science & environmental epidemiology*, *29*(4), 447-456.

[101]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

[102]. Lee, D. S., & Shan, J. (2003). Combining LIDAR elevation data and IKONOS multispectral imagery for coastal classification mapping. *Marine Geodesy*, *26*(1-2), 117-127.

[103]. Li, J., Qin, Q., Han, J., Tang, L. A., & Lei, K. H. (2015). Mining trajectory data and geotagged data in social media for road map inference. *Transactions in GIS*, *19*(1), 1-18.

[104]. Liang, H., Sun, X., Sun, Y., & Gao, Y. (2017). Text feature extraction based on deep learning: a review. *EURASIP journal on wireless communications and networking*, *2017*(1), 1-12.

[105]. Lin, H. W., Tegmark, M., & Rolnick, D. (2017). Why does deep and cheap learning work so well?. *Journal of Statistical Physics*, *168*(6), 1223-1247.

[106]. Lin, Y. Z., Nie, Z. H., & Ma, H. W. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering*, *32*(12), 1025-1046.

[107]. Liu, L., Coops, N. C., Aven, N. W., & Pang, Y. (2017). Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. *Remote Sensing of Environment*, *200*, 170-182.

[108]. Liu, X., de Sherbinin, A., & Zhan, Y. (2019). Mapping urban extent at large spatial scales using machine learning methods with VIIRS Nighttime light and MODIS daytime NDVI data. *Remote Sensing*, *11*(10), 1247.

[109]. Lo, C. S., & Wang, C. M. (2012). Support vector machine for breast MR image classification. *Computers & Mathematics with Applications*, *64*(5), 1153-1162.

[110]. Lodha, S. K., Kreps, E. J., Helmbold, D. P., & Fitzpatrick, D. (2006, June). Aerial LiDAR data classification using support vector machines (SVM). In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)* (pp. 567-574). IEEE.

[111]. Long, Y., Gong, Y., Xiao, Z., & Liu, Q. (2017). Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(5), 2486-2498.

[112]. Lopez-Torrijos, R. (2018). *Best Practices in Hydrography Extraction from Lidar: Status Update, HydroShare,* http://www.hydroshare.org/resource/7ec2a044ccf24482b1527858df669e19.

[113]. Lukas, V., Eldridge, D.F., Jason, A.L., Saghy, D.L., Steigerwald, P.R., Stoker, J.M., Sugarbaker, L.J., & Thunen, D.R. (2015). *Status report for the 3D Elevation Program, 2013–2014: U.S. Geological Survey Open-File Report 17, 2015–1161,* http://dx.doi.org/10.3133/ofr20151161.

[114]. Ma, L., Fu, T., Blaschke, T., Li, M., Tiede, D., Zhou, Z., Ma, X., & Chen, D. (2017). Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. *ISPRS International Journal of Geo-Information*, *6*(2), 51.

[115]. Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). High-resolution aerial image labeling with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(12), 7092-7103.

[116]. Maidment, D. R. (2016). Conceptual framework for the national flood interoperability experiment. *Journal of The American Water Resources Association, 53*(2), 245–257.

[117]. Maidment, D. R., & Morehouse, S. (2002). *Arc Hydro: GIS for water resources*. ESRI, Inc..

[118]. Maturana, D., & Scherer, S. (2015a, May). 3d convolutional neural networks for landing zone detection from lidar. In *2015 IEEE international conference on robotics and automation (ICRA)* (pp. 3471-3478). IEEE.

[119]. Maturana, D., & Scherer, S. (2015b, September). Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 922-928). IEEE.

[120]. McDermid, G. J., Hall, R. J., Sanchez-Azofeifa, G. A., Franklin, S. E., Stenhouse, G. B., Kobliuk, T., & LeDrew, E. F. (2009). Remote sensing and forest inventory for wildlife habitat assessment. *Forest Ecology and Management*, *257*(11), 2262-2269.

[121]. McManamon, P. F., Banks, P. S., Beck, J. D., Fried, D. G., Huntington, A. S., & Watson, E. A. (2017). Comparison of flash lidar detector options. *Optical Engineering*, *56*(3), 031223.

[122]. Metz, M., Mitasova, H., & Harmon, R. S. (2011). *Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search*. ULM UNIV (GERMANY FR).

[123]. Mnih, V., & Hinton, G. E. (2010, September). Learning to detect roads in high-resolution aerial images. In *European Conference on Computer Vision* (pp. 210-223). Springer, Berlin, Heidelberg.

[124]. Mo, H., Luo, C., & Liu, K. (2016, June). Robot indoor navigation based on computer vision and machine learning. In *International Conference on Swarm Intelligence* (pp. 528-534). Springer, Cham.

[125]. Moenning, C., & Dodgson, N. A. (2003). *Fast marching farthest point sampling* (No. UCAM-CL-TR-562). University of Cambridge, Computer Laboratory.

[126]. Montgomery, D. R., & Foufoula-Georgiou, E. (1993). Channel network source representation using digital elevation models. *Water Resources Research*, *29*(12), 3925-3934.

[127]. Morsy, S., Shaker, A., El-Rabbany, A., & LaRocque, P. E. (2016). Airborne multispectral LiDAR data for land-cover classification and land/water mapping using different spectral indexes. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 3*(3).

[128]. Muller, R. A., & Oberlander, T. M. (1978). *Physical Geography Today: Portrait of a Planet*. Random House, New York, New York. 590pp.

[129]. Nadeau, T., & Rains, M.C. (2007). Hydrological connectivity between headwater streams and downstream waters: how science can inform policy, *Journal of the American Water Resources Association (JAWRA), 43*(1):118-133.

[130]. Nguyen, M. Q., Atkinson, P. M., & Lewis, H. G. (2005). Superresolution mapping using a Hopfield neural network with LIDAR data. *IEEE Geoscience and Remote Sensing Letters, 2*(3), 366-370.

[131]. NOAA (National Oceanic and Atmospheric Administration). National Water Model: Improving NOAA's Water Prediction Services. (2016).

[132]. Nogueira, J. M., Ruffault, J., Chuvieco, E., & Mouillot, F. (2017). Can we go beyond burned area in the assessment of global remote sensing products with fire patch metrics? *Remote Sensing*, *9*(1), 7.

[133]. O'Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer vision, graphics, and image processing*, *28*(3), 323-344.

[134]. Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

[135]. Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1717-1724).

[136]. Ou, C., Yang, J., Du, Z., Zhang, X., & Zhu, D. (2019). Integrating Cellular Automata with Unsupervised Deep-Learning Algorithms: A Case Study of Urban-Sprawl Simulation in the Jingjintang Urban Agglomeration, China. *Sustainability*, *11*(9), 2464.

[137]. Padmanabhan, A., Wang, S., Cao, G., Hwang, M., Zhang, Z., Gao, Y., Soltani, K., & Liu, Y.Y. (2014). FluMapper: A cyberGIS application for interactive analysis of massive location-based social media. *Concurrency and Computation: Practice and Experience*, *26*(13), 2253-2265.

[138]. Papon, J., Abramov, A., Schoeler, M., & Worgotter, F. (2013). Voxel cloud connectivity segmentation-supervoxels for point clouds. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2027-2034).

[139]. Passalacqua, P., Belmont, P., & Foufoula-Georgiou, E. (2012). Automatic geomorphic feature extraction from lidar in flat and engineered landscapes. *Water Resources Research*, *48*(3).

[140]. Passalacqua, P., Do Trung, T., Foufoula-Georgiou, E., Sapiro, G., & Dietrich, W. E. (2010). A geometric framework for channel network extraction from lidar: Nonlinear diffusion and geodesic paths. *Journal of Geophysical Research: Earth Surface*, *115*(F1).

[141]. Paul, R., Arkin, J., Roy, N., & M Howard, T. (2016). Efficient grounding of abstract spatial concepts for natural language interaction with robot manipulators.

[142]. Poghosyan, A. (2018). Quantifying urban growth in 10 post-Soviet cities using Landsat data and machine learning. *International Journal of Remote Sensing*, *39*(23), 8688-8702.

[143]. Popescu, S. C., Zhao, K., Neuenschwander, A., & Lin, C. (2011). Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sensing of Environment*, *115*(11), 2786-2797.

[144]. Poppenga, S. K., Gesch, D. B., & Worstell, B. B. (2013). Hydrography Change Detection: The Usefulness of Surface Channels Derived From LiDAR DEMs for Updating Mapped Hydrography 1. *JAWRA Journal of the American Water Resources Association*, *49*(2), 371-389.

[145]. Prokhorov, D. (2010). A convolutional learning system for object classification in 3-D lidar data. *IEEE Transactions on neural networks*, *21*(5), 858-863.

[146]. Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017a). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 652-660).

[147]. Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017b). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems* (pp. 5099-5108).

[148]. Qin, R., Tian, J., & Reinartz, P. (2016). 3D change detection–approaches and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, *122*, 41-56.

[149]. Rahimi, A., & Recht, B. (2008). Random features for large-scale kernel machines. In *Advances in neural information processing systems* (pp. 1177-1184).

[150]. Rawat, J. S., & Kumar, M. (2015). Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. *The Egyptian Journal of Remote Sensing and Space Science*, *18*(1), 77-84.

[151]. Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 806-813).

[152]. Reiche, J., Lucas, R., Mitchell, A. L., Verbesselt, J., Hoekman, D. H., Haarpaintner, J., Kellndorfer, J.M., Rosenqvist, A., Lehmann, E.A., Woodcock, C.E., Seifert, F.M., & Herold, M. (2016). Combining satellite data for better tropical forest monitoring. *Nature Climate Change*, *6*, 120-122.

[153]. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195-204.

[154]. Resta, S., Acito, N., Diani, M., Corsini, G., Opsahl, T., & Haavardsholm, T. V. (2011, July). Detection of small changes in airborne hyperspectral imagery: Experimental results

over urban areas. In *2011 6th International Workshop on the Analysis of Multi-temporal Remote Sensing Images (Multi-Temp)* (pp. 5-8). IEEE.

[155]. Reymann, C., & Lacroix, S. (2015, September). Improving LiDAR point cloud classification using intensities and multiple echoes. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 5122-5128). IEEE.

[156]. Richter, R., Behrens, M., & Döllner, J. (2013). Object class segmentation of massive 3D point clouds of urban areas using point cloud topology. *International Journal of Remote Sensing*, *34*(23), 8408-8424.

[157]. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

[158]. Russell, P. (2008). Mapping headwater streams: intermittent and perennial headwater stream model development and spatial application. *Final Report for Federal Highway Administration Contract Feasibility Study WBS: 36486.4.2*. January 28. Raleigh: North Carolina Division of Water Quality.

[159]. Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach. Malaysia*. Pearson Education Limited.

[160]. Rusu, R. B., & Cousins, S. (2011, May). 3d is here: Point cloud library (pcl). In *2011 IEEE international conference on robotics and automation* (pp. 1-4). IEEE.

[161]. Salamah, A. H., Tamazin, M., Sharkas, M. A., & Khedr, M. (2016, October). An enhanced WiFi indoor localization system based on machine learning. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (pp. 1-8). IEEE.

[162]. Sangireddy, H., Stark, C. P., Kladzyk, A., & Passalacqua, P. (2016). GeoNet: An open source software for the automatic and objective extraction of channel heads, channel network, and channel morphology from high resolution topography data. *Environmental Modelling & Software*, *83*, 58-73.

[163]. Santosuosso, U., & Papini, A. (2018). Geo-profiling: beyond the current limits. A preliminary study of mathematical methods to improve the monitoring of invasive species. *Russian Journal of Ecology*, *49*(4), 362-370.

[164]. Sasaki, T., Imanishi, J., Ioki, K., Morimoto, Y., & Kitada, K. 2012. Object-based classification of land cover and tree species by integrating airborne LiDAR and high spatial resolution imagery data. *Landscape and Ecological Engineering, 8*(2), 157-171.

[165]. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

[166]. Schlkopf, B., Smola, A. J., & Bach, F. (2018). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. the MIT Press.

[167]. Schultz, L. D., Heck, M. P., Hockman-Wert, D., Allai, T., Wenger, S., Cook, N. A., & Dunham, J. B. (2017). Spatial and temporal variability in the effects of wildfire and drought on thermal habitat for a desert trout. *Journal of Arid Environments*, *145*, 60-68.

[168]. Schwarz, B. (2010). LIDAR: Mapping the world in 3D. *Nature Photonics, 4*(7), 429.

[169]. Scott, R. A. (2017). *The making of blind men*. Routledge.

[170]. Sharma, R., Xu, Z., Sugumaran, R., & Oliveira, S. (2016). Parallel landscape driven data reduction & spatial interpolation algorithm for big LiDAR data. *ISPRS International Journal of Geo-Information, 5*(6), 97.

[171]. Shavers, E., & Stanislawski, L. (2018). Streams do work: measuring the work of low-order streams on the landscape using point clouds. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.*

[172]. Shen, H., Huang, L., Zhang, L., Wu, P., & Zeng, C. (2016). Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: A 26-year case study of the city of Wuhan in China. *Remote Sensing of Environment*, *172*, 109-125.

[173]. Sheng, J., Wilson, J. P., Chen, N., Devinny, J. S., & Sayre, J. M. (2007). Evaluating the quality of the National Hydrography Dataset for watershed assessments in metropolitan regions. *GIScience & Remote Sensing*, *44*(3), 283-304.

[174]. Shiklomanov, I. (1993). World fresh water resources. In *Gleick, P.H. (editor) Water in Crisis: A Guide to the World's Fresh Water Resources*. Oxford University Press, New York.

[175]. Simley, J. D., & Carswell Jr, W. J. (2009). The national map—Hydrography. *US Geological Survey Fact Sheet*, *3054*(4).

[176]. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *Proceedings of the International Conference on Learning Representations*, San Diego, CA, USA, 7–9 May 2015.

[177]. Singh, K. K., Vogler, J. B., Shoemaker, D. A., & Meentemeyer, R. K. (2012). LiDAR-Landsat data fusion for large-area assessment of urban land cover: Balancing spatial resolution, data volume and mapping accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing, 74*, 110-121.

[178]. Singla, S. K., Garg, R. D., & Dubey, O. P. (2018). Spatiotemporal analysis of LANDSAT Data for Crop Yield Prediction. *Journal of Engineering Science & Technology Review, 11*(3).

[179]. Sinha, S., Jeganathan, C., Sharma, L. K., & Nathawat, M. S. (2015). A review of radar remote sensing for biomass estimation. *International Journal of Environmental Science and Technology, 12*(5), 1779-1792.

[180]. Slater, P. N. (1980). Remote sensing: optics and optical systems. Reading, Mass., Addison-Wesley Publishing Co.(Remote Sensing. No. 1), 1980. 593 p., 1.

[181]. Socher, R., Huval, B., Bath, B., Manning, C. D., & Ng, A. Y. (2012). Convolutional-recursive deep learning for 3d object classification. In *Advances in neural information processing systems* (pp. 656-664).

[182]. Song, X. P., Sexton, J. O., Huang, C., Channan, S., & Townshend, J. R. (2016). Characterizing the magnitude, timing and duration of urban growth from time series of Landsat-based estimates of impervious cover. *Remote Sensing of Environment*, *175*, 1-13.

[183]. Stanislawski, L. V. (2009). Feature pruning by upstream drainage area to support automated generalization of the United States National Hydrography Dataset. *Computers, Environment and Urban Systems*, *33*(5), 325-333.

[184]. Stanislawski, L. V., Survila, K., Wendel, J., Liu, Y., & Buttenfield, B. P. (2018). An open source high-performance solution to extract surface water drainage networks from diverse terrain conditions. *Cartography and Geographic Information Science*, *45*(4), 319-328.

[185]. Stoker, J. M., Abdullah, Q. A., Nayegandhi, A., & Winehouse, J. (2016). Evaluation of single photon and Geiger mode LiDAR for the 3D elevation program. *Remote Sensing*, *8*(9), 767.

[186]. Stow, D. A., & Chen, D. M. (2002). Sensitivity of multitemporal NOAA AVHRR data of an urbanizing region to land-use/land-cover changes and misregistration. *Remote Sensing of Environment*, *80*(2), 297-307.

[187]. Sun, Y., Zhang, X., Xin, Q., & Huang, J. (2018). Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data. *ISPRS journal of photogrammetry and remote sensing, 143*, 3-14.

[188]. Syed, S., Dare, P., & Jones, S. (2005, September). Automatic classification of land cover features with high resolution imagery and lidar data: an object-oriented approach. In *Proceedings of SSC2005 spatial intelligence, innovation and praxis: the national biennial conference of the spatial sciences institute* (pp. 512-522). Spatial Science Institute Melbourne: Melbourne.

[189]. Tarboton, D. G., Bras, R. L., & Rodriguez-Iturbe, I. (1991). On the extraction of channel networks from digital elevation data. *Hydrological processes*, *5*(1), 81-100.

[190]. Teng, Y., Bi, D., Guo, X., & Paul, R. (2018). Predicting the Epidemic Potential and Global Diffusion of Mosquito-Borne Diseases Using Machine Learning. *Available at SSRN 3260785*.

[191]. Teo, T. A., & Shih, T. Y. (2013). Lidar-based change detection and change-type determination in urban areas. *International Journal of Remote Sensing*, *34*(3), 968-981.

[192]. Terziotti, S., Adkins, K., Aichele, S., Anderson, R., & Archuleta, C. (2018). Testing the waters: integrating hydrography and elevation in national hydrography mapping. *AWRA Water Resources IMPACT, 20*(2), 28-29.

[193]. Thakuriah, P. V., Tilahun, N., & Zellner, M. (Eds.). (2016). *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics*. Springer.

[194]. Toth, C. K., & Paska, E. (2007). Mobile mapping and autonomous vehicle navigation. Revue Francaise Photogramm. *Teledetection, 185*, 57-61.

[195]. Toth, C. K., Koppanyi, Z., & Lenzano, M. G. (2018). New source of geospatial data: crowdsensing by assisted and autonomous vehicle technologies. International Archives of the Photogrammetry, *Remote Sensing & Spatial Information Sciences, 42*.

[196]. Tuia, D., Kellenberger, B., Pérez-Suey, A., & Camps-Valls, G. (2018, July). A deep network approach to multitemporal cloud detection. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 4351-4354). IEEE.

[197]. Valero, S., Chanussot, J., Benediktsson, J. A., Talbot, H., & Waske, B. (2010). Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images. *Pattern Recognition Letters*, *31*(10), 1120-1127.

[198]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).

[199]. Venerandi, A., Zanella, M., Romice, O., Dibble, J., & Porta, S. (2017). Form and urban change–An urban morphometric study of five gentrified neighbourhoods in London. *Environment and Planning B: Urban Analytics and City Science*, *44*(6), 1056-1076.

[200]. VoPham, T., Hart, J. E., Laden, F., & Chiang, Y. Y. (2018). Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. *Environmental Health*, *17*(1), 40.

[201]. Wang, B., Choi, J., Choi, S., Lee, S., Wu, P., & Gao, Y. (2017). Image fusion-based land cover change detection using multi-temporal high-resolution satellite images. *Remote Sensing, 9*(8), 804.

[202]. Wang, B., Choi, S., Byun, Y., Lee, S., & Choi, J. (2015). Object-based change detection of very high resolution satellite imagery using the cross-sharpening of multitemporal data. *IEEE Geoscience and Remote Sensing Letters, 12*(5), 1151-1155.

[203]. Wang, D., Wan, B., Qiu, P., Su, Y., Guo, Q., & Wu, X. (2018). Artificial mangrove species mapping using pléiades-1: An evaluation of pixel-based and object-based classifications with selected machine learning algorithms. *Remote Sensing, 10*(2), 294.

[204]. Wang, S. (2010). A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers, 100*(3), 535-557.

[205]. Wang, S. (2016). CyberGIS and spatial data science. *GeoJournal, 81*(6), 965-968.

[206]. Wang, S., & Goodchild, M.F. (2019). *CyberGIS for Geospatial Innovation and Discovery*. Springer, Dordrecht, Netherlands, DOI: 10.1007/978-94-024-1531-5.

[207]. Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M. F., Liu, Y., & Nyerges, T. L. (2013). CyberGIS software: a synthetic review and integration roadmap. *International Journal of Geographical Information Science, 27*(11), 2122-2145.

[208]. Wang, S., Liu, Y., & Padmanabhan, A. (2016). Open cyberGIS software for geospatial research and education in the big data era. *SoftwareX, 5*: 1-5.

[209]. Wang, X., Gong, W., Huang, X., Liu, T., Zhou, Y., & Li, H. (2018). Assessment of Eco-environmental Quality on Land Use and Land Cover Changes Using Remote Sensing and GIS: A Case Study of Miyun County. *Nature Environment and Pollution Technology, 17*(3), 739-746.

[210]. Wen, Z., Shi, J., Li, Q., He, B., & Chen, J. (2018). ThunderSVM: A fast SVM library on GPUs and CPUs. *The Journal of Machine Learning Research, 19*(1), 797-801.

[211]. Wright, W., Nielsen, B., Mullen, J. D., & Dowd, J. (2012). *Agricultural groundwater policy during drought: a spatially differentiated approach for the Flint River Basin* (No. 323-2016-11832).

[212]. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J. (2015). 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1912-1920).

[213]. Wulder, M. A., Han, T., White, J. C., Sweda, T., & Tsuzuki, H. (2007). Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization. *Remote Sensing of Environment*, *110*(1), 123-137.

[214]. Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E.A., Shaw, J., Masek, J.G., & Roy, D. P. (2016). The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, *185*, 271-283.

[215]. Xiao, W., Vallet, B., Brédif, M., & Paparoditis, N. (2015). Street environment change detection from mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, *107*, 38-49.

[216]. Xu, H., Cheng, L., Li, M., Chen, Y., & Zhong, L. (2015). Using octrees to detect changes to buildings and trees in the urban environment from airborne LiDAR data. *Remote Sensing*, *7*(8), 9682-9704.

[217]. Xu, Z., Guan, K., Casler, N., Peng, B., & Wang, S. (2018). A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal Landsat imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *144*, 423-434.

[218]. Xu, Z., Mountrakis, G., & Quackenbush, L. J. (2017). Impervious surface extraction in imbalanced datasets: integrating partial results and multi-temporal information in an iterative one-class classifier. *International journal of remote sensing*, *38*(1), 43-63.

[219]. Yan, W. Y., Shaker, A., & El-Ashmawy, N. (2015). Urban land cover classification using airborne LiDAR data: A review. *Remote Sensing of Environment*, *158*, 295-310.

[220]. Yin, D., Liu, Y., Hu, H., Terstriep, J., Hong, X., Padmanabhan, A., & Wang, S. (2019). CyberGIS-Jupyter for reproducible and scalable geospatial analytics. *Concurrency and Computation: Practice and Experience*, *31*(11), e5040.

[221]. Yousefhussien, M., Kelbe, D. J., Ientilucci, E. J., & Salvaggio, C. (2018). A multi-scale fully convolutional network for semantic labeling of 3D point clouds. *ISPRS journal of photogrammetry and remote sensing*, *143*, 191-204.

[222]. Yuan, F., Sawaya, K. E., Loeffelholz, B. C., & Bauer, M. E. (2005). Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. *Remote sensing of Environment*, *98*(2-3), 317-328.

[223]. Zelener, A., & Stamos, I. (2016, October). CNN-based object segmentation in urban lidar with missing points. In *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 417-425). IEEE.

[224]. Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 22-40.

[225]. Zhang, Z. (2018, June). Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)* (pp. 1-2). IEEE.

[226]. Zhang, Z., Hu, H., Yin, D., Kashem, S., Li, R., Cai, H., Perkins, D., and Wang, S. (2019). A cyberGIS-enabled multi-criteria spatial decision support system: A case study on flood emergency management. *International journal of digital earth*, 12(11), 1364-1381.

[227]. Zhao, Y., Feng, D., Yu, L., Wang, X., Chen, Y., Bai, Y., Hernández, H.J., Galleguillos, M., Estades, C., Biging, G.S., Radke, J.D., & Gong, P. (2016). Detailed dynamic land cover mapping of Chile: Accuracy improvement by integrating multi-temporal data. *Remote Sensing of Environment, 183*, 170-185.

[228]. Zhou, W. (2013). An object-based approach for urban land cover classification: integrating LiDAR height and intensity data. *IEEE Geoscience and Remote Sensing Letters, 10*(4), 928-931.

[229]. Zhou, X., Zheng, H. B., Xu, X. Q., He, J. Y., Ge, X. K., Yao, X., Cheng, T., Zhu, Y., Cao, W.X., & Tian, Y. C. (2017). Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 130*, 246-255.

[230]. Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine, 5*(4), 8-36.

[231]. Zhu, X., & Toutin, T. (2011, July). A multi-window texture classification and object-oriented feature extraction method with airborne lidar products. In *2011 IEEE International Geoscience and Remote Sensing Symposium* (pp. 3370-3373). IEEE.

[232]. Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote sensing of Environment*, *144*, 152-171.